

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number
WO 02/102841 A2

(51) International Patent Classification⁷: **C07K 14/00**

KISS, Péter [HU/HU]; Földmíves u. 1., H-6725 Szeged (HU).

(21) International Application Number: PCT/HU02/00055

(74) Agent: **SZENTPÉTERI, Ádám**; S.B.G. & K. Patent and Law Offices, Andrásy út 113., H-1060 Budapest (HU).

(22) International Filing Date: 19 June 2002 (19.06.2002)

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AU, BA, BB, BG, BR, BZ, CA, CN, CO, CR, CU, CZ, DM, DZ, EC, EE, GD, GE, HR, ID, IL, IN, IS, JP, KP, KR, LC, LK, LR, LT, LV, MA, MG, MK, MN, MX, MZ, NO, NZ, OM, PH, PL, RO, SG, SI, SK, SL, TN, TR, TT, TZ, UA, US, UZ, VN, YU, ZA, ZM.

(26) Publication Language: English

(30) Priority Data:
P0102561 20 June 2001 (20.06.2001) HU

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): **MTA SZEGEDI BIOLÓGIAI KÖZPONT** [HU/HU]; Genetikai Intézete, Temesvári krt 62., H-6726 Szeged (HU).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **KISS, György, Botond** [HU/HU]; Szabad Sajtó ú. 48., H-6725 Szeged (HU). **ENDRE, Gabriella** [HU/HU]; Budapesti krt. 2/A, H-6723 Szeged (HU). **KERESZT, Attila** [HU/HU]; Zöldfa u. 4., H-6723 Szeged (HU). **KEVEI, Zoltán** [HU/HU]; Körtöltés u. 52., H-6724 Szeged (HU). **KALÓ, Péter** [HU/HU]; Retek u. 19/b, H-6723 Szeged (HU).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IDENTIFICATION AND CLONING OF THE RECEPTOR GENE FOR SYMBIOTIC NITROGEN FIXATION

(57) Abstract: The invention relates to the *Medicago sativa* DNA sequence encoding the NORK polypeptide as well as to the protein determined by this DNA sequence. In another aspect, the invention relates to conserved DNA regions, primers, DNA probes, polypeptides and corresponding antibodies characteristic to the NORK gene family, as well as to methods and their utilization to isolate additional NORK genes and proteins from different legumes or other plants, and their use for producing transgenic plants. In another aspect, the invention relates to a method of cloning a DNA sequence encoding a function involved in one of the steps of symbiotic nitrogen fixation, as well as to methods of producing transgenic plant cells and cloned entire plants wherein the biosynthesis of the NORK polypeptide takes place. The invention also relates to the identification of a gene present in leguminous plants, encoding a function indispensable for symbiotic nitrogen fixation, to the sequencing of the gene, confirmation of the biological function of the gene, and the transformation of the gene into plants unable for symbiotic nitrogen fixation.



WO 02/102841 A2

Identification and cloning of the receptor gene for symbiotic nitrogen fixation

TECHNICAL FIELD

5

The invention relates to the *Medicago sativa* DNA sequence encoding the NORK polypeptide as well as to the protein determined by this DNA sequence. In another aspect, the invention relates to conserved DNA regions, primers, DNA probes, polypeptides and corresponding antibodies characteristic to the NORK gene family, as well as to methods and their utilization to isolate additional NORK genes and proteins from different legumes or other plants, and their use for producing transgenic plants.

In another aspect, the invention relates to a method of cloning a DNA sequence encoding a function involved in one of the steps of symbiotic nitrogen fixation, as well as to methods of producing transgenic plant cells and cloned entire plants wherein the biosynthesis of the NORK polypeptide takes place.

The invention also relates to the identification of a gene present in leguminous plants, encoding a function indispensable for symbiotic nitrogen fixation, to the sequencing of the gene, confirmation of the biological function of the gene, and the transformation of the gene into plants unable for symbiotic nitrogen fixation.

25 BACKGROUND ART

It is well known that symbiotic nitrogen fixation covers the requirements of the nitrogen demands of a plant and ensures normal growth, development and seed production in soils with limited amount of fixed nitrogen without added artificial nitrogen fertilizers (1). Plants capable of symbiotic fixation differ from other plants in that they carry all of the genetic determinants in their genome which are needed for the biological process of symbiotic nitrogen fixation. For those experts

who are familiar with the recombinant DNA techniques it is obvious that these genes can be identified, isolated and transferred to other plants on such a way that they retain their function. In theory, all genes required for symbiotic nitrogen fixation can be transferred to recipient plants which eventually will be able to
5 establish symbiotic nitrogen fixation, consequently, they can be grown on a more economical and environmental friendly manner.

The identification of genes involved in symbiotic nitrogen fixation are being studied very intensively worldwide, still the nucleotide and amino acid sequences
10 which are presented in this patent application are completely novel; we had access only to randomly sequenced cDNA sequences.

The symbiotic nitrogen fixation is a biological process established between particular dicotyledonous plants and soil bacteria (2, 3). The symbiotic nitrogen
15 fixation takes place mostly in the root system where root nodules develop, and are invaded by the bacteria. During the development of root nodules bacteria enter the plant cells and change to bacteroids, and in such form they fix nitrogen that is they convert the inert dinitrogen of the air into ammonia. This ammonia (which is a form of the so-called fixed nitrogen) can be utilized by the plants for their development.
20 Since no artificial nitrogen fertilization is needed, symbiotic nitrogen fixation is an environmental friendly, cost effective and improve the quality of the soil, therefore, these plants are very suitable as green crops. Agricultural legumes, which are capable to fix the nitrogen from the air, are therefore very important part of the agriculture. Their seeds and green mass contribute to large extent to the human
25 and animal amino acid and oil supply, therefore, legumes are extremely important in the worldwide food supply.

Some soil bacteria, like *Rhizobia* and *Actinomycetes*, are able to establish symbiotic nitrogen fixation relationship with given plants, like many legumes,
30 *Trema*, *Casuarina*, *Alnus*, *Parasponia* (2, 3).

As a consequence of the evolution this symbiotic relationship is strictly species specific, a given group of bacteria establishes symbiosis only with a given group of plants (2, 3, 4, 5). Such symbiosis is established between *Sinorhizobium meliloti* and *Medicago*, *Melilotus*, *Trigonella* species (6), between *Rhizobium leguminosarum* bv. *viciae* and *Vicia* species, between *Rhizobium leguminosarum* bv. *trifolii* and given members of *Trifolium* species, between *Rhizobium etli* and *Phaseolus* species, between *Bradyrhizobium japonicum* and *Glycine max*, between *Azorhizobium caulinodans* and *Sesbania rostrata*, between *Mesorhizobium loti* and *Lotus japonicus* (7).

10

The biogenesis and development of the symbiotic nodule, as well as the invasion of the plants cells by the bacteria is a complex, multi-step process taking place through series of gene-interaction involving both bacterial and plant symbiotic genes. The formation of the nodule, which is inevitable for the symbiotic nitrogen fixation according to our present knowledge, starts to develop after a specific signal exchange between the microsymbiont (the bacteria) and the macrosymbiont (the plants). Through the root system, the macrosymbionts excrete specific flavonoid and other molecules into the environment, which are characteristic to and determined by the genetic material of the plant species (8, 9).

As a response to these excreted molecules compatible bacteria approach by chemotaxis the plant root system (10) and start to produce a signal molecule, the so-called lipo-chitooligosaccharides abbreviated as Nod factors, which in turn are specific nodulation signal for the plant (11, 12, 13). During nodule formation compatible bacteria attach to the root-hair surface on a polar manner. In this joining special molecules like bacterial pili, plant lectins and other adhesive molecules are active components (14). Upon the action of the symbiotic specific Nod factors, characteristic plant responses occur like root-hair deformation, curling and branching (15). Inside the root tissue, the cortical cells start to divide. As a response of the specific Nod factor produced by the compatible bacteria a symbiotic nodule start to develop on the root or on the stem depending on the type of symbiosis. Some of the specific Nod factors are shown in Fig. 1. without the sake of completeness (5, 7). As a consequence of the root-hair curling

30

provoked by the specific Nod factor, the bacteria become surrounded and enter the so-called infection threads produced by the plant. The infection threads start to grow towards the core tissue of the plant, where they start to branch. Inside the infection thread, the bacteria divide and move along the infection threads. At the
5 same time the cortical cells continue to divide and the nodule meristeme is formed. The branching infection threads enter the dividing cells and the bacteria are released inside the plants cells, where they continue to divide and in a short time they occupy most of the plant cytoplasm establishing the so-called infected cells. In the infected cells bacteria are surrounded by plant plasma membrane
10 (peribacteroid membrane) in which the bacteria change to bacteroids, in which form they are capable to fix nitrogen (16, 17). The derepression of the nitrogenase enzyme complex takes place only in the bacteroids. The bacteroids, with the help of their nitrogenase enzyme complex, can reduce the inert dinitrogen into ammonia, the so-called combined nitrogen, which can be utilized by the
15 macrosymbiont. Most of the ammonia formed during nitrogen fixation (that part which is not utilized by the bacteroids) is excreted by the bacteroids. The excreted ammonia is taken up by the nodule cells and transferred towards the vascular tissue by a proper conducting mechanism. The fully developed and active nodules are able to supply the plants with enough combined nitrogen which is needed for
20 their normal growth and development.

Root-hair deformation and cortical cell division can be evoked by very low (10^{-8} - 10^{-12} M) concentration of the compatible Nod factor even if the bacteria are not present, if the Nod factor is applied to the root. Following Nod factor treatment
25 normal nodule structures are formed on the plant roots, in which no bacteria are present. From this phenomenon it is obvious, that the Nod factors are recognized by the root cells, which in turn start the specific signal transduction pathway leading to nodule formation. The recognition of the excreted Nod factor on the surface of the root cells is taking place by an unknown mechanism. Nod factor
30 binding receptors were looked for by biochemical methods, but no such a supposed protein responsible for the binding and transmission of the Nod factor could have been identified so far (18). However, it is generally accepted, that the

bacterial Nod factor is recognized and bound by a specific receptor and through this mechanism the signal transduction cascade evokes nodule induction and development.

5 The steps leading to nodule development, the functions involved, and the genes coding for these functions in the plants are not known at present, however, one can understand that the presence of the microsymbiont for the formation of the nodule structure is not necessary. Some plants like alfalfa (*Medicago sativa* L.) is able to form spontaneous nodules. Under sterile conditions and in the
10 absence of microbes and higher concentration of combined nitrogen (more than 1 mM) alfalfa develops some nodule-like structure, which are very similar to the morphology and tissue characteristics of the normal nodules (19). As a consequence, all those genes and functions, which are necessary for nodule development are encoded by the plant genetic material. However, the Nod factors
15 produced by the compatible microsymbiont are responsible to what extend, and what type of nodules are formed by the plant.

 Several genes, both bacterial and plant origin, are involved in the development and effective functioning of the nodules. Most of the genes in *S.*
20 *Meliloti* and in other *Rhizobia* are known (5, 20, and personal information from Michael Göttfert, Technische Universität Dresden, Germany). According to Prof. Alfred Pühler, University of Bielefeld, Germany the nucleotide sequence of the entire *S. meliloti* genome is known. However, the identification and the function of the plant genes involved in this biological process is still in its infancy. At present,
25 biochemical and combined genetic and biochemical approaches are applied to study plant genes. Genes identified by biochemical methods are questionable as to whether they are involved in symbiotic nitrogen fixation. The nodule specific expression of many genes isolated by biochemical approaches was demonstrated nevertheless their real function was not uncovered (21, 22, 23), or it turned out
30 that they are not essential to effective symbiotic nitrogen fixation (24).

Among others the *Medicago* EST programs will be useful to get acquainted with the nodule specific genes. These programs generated until now more than 100 000 partial EST sequences. These sequences are open for public on the Internet (<http://www.ncbi.nlm.nih.gov>). Among these sequences there are three partial cDNA sequences homologous to the NORK gene of *M. truncatula*. Form the mere sequence, however, it is impossible to predict the function of the gene, therefore more analysis is needed to reveal the function. The most powerful tool for the study of gene functions if mutant alleles of the genes are available. One can generate a mutant allele of a gene, or one can isolate a gene from a mutant allele. The first possibility (homologue recombination) is not available for plants. The later, however, is feasible through map based cloning or insertion mutagenesis (transposons, T-DNA) based gene isolation procedures. According to mutant phenotype (T-DNA insertion) only one gene has been isolated form *Lotus japonicus*: this gene (*nin*) encoded a transcription factor (25). In this patent we describe for the first time the isolation of a gene by map based cloning that is the isolation of the NORK gene with the help of the *Medicago sativa* mutant MN1008.

Classical genetic strategy leads to unambiguous results. The isolated gene is an essential component of the biological process in question if the gene affected by the mutation has a phenotype (unable to carry out certain functions or less active in a given function) and the gene isolated based on the phenotype. In the mutant individuals an essential function affected which is involved in the biological process studied. Appropriate methods are available by which the mutant gene ant ist wild type counterpart can be isolated. In order to acquire mutant individuals mutations can be induced by physical, chemical or biological means. Mutant genes, in which no insertions of known sequences are present, or the nature of the mutation is not known (like in the case of large deletions, inversions) can be isolated by the so-called map based cloning. The principal of the map based cloning is the following: first the phenotype is mapped genetically with molecular (DNA) markers, and then with the help of these markers chromosomal walking is performed and the gene is isolated form clones originated form large insert libraries (26).

Mutant plants impaired in symbiotic nitrogen fixation can be grouped into two categories: mutants in the first class are the Nod⁻ mutants that is no nodules appear on their root system; mutants in the second class are the Fix⁻ mutants which develop nodules or nodule like structures but the nodules are ineffective, that is no effective nitrogen fixation takes place. It is a characteristic phenomenon of both mutant types, that in the presence of the compatible bacteria and in the absence of combined nitrogen the symptoms of nitrogen deficiency develop on the mutants, that is yellow leaves and stunted growth. There are many symbiotic mutants available published in the literature. Among the Nod⁻ mutants the most important ones are the *Medicago sativa* MN1008 mutant (27, 28), the *Medicago truncatula* B129, Tr25, Trv25, B85 mutants (29), the *Pisum sativum* sym8, sym10, sym19 mutants (30), the *Lotus japonicus* sym2 mutant (31), and the *Glycine max* nod49 and nod139 mutants (14).

15

The MnNC-1008 (NN) alfalfa (*Medicago sativa*) mutant (abbreviated as MN1008) is incapable to form nodules on its root system in the presence of its compatible microsymbiont (*Sinorhizobium meliloti*) (27, 28). Further studies showed that there was no sign of any response detectable by microscopy, which was characteristic to the wild type plant early nodulation response. There was no root hair deformation and cortical cell division (32), in addition to the lack of the specific calcium spiking response (special calcium oscillation inside the cells) (33). According to unpublished results, there was no root hair curling or cortical cell division observed in the mutant MN1008 (unlike in the wild type plant) upon addition the specific *S. meliloti* Nod factor. Mutant MN1008, on the other hand, was able to develop nodules, the so-called spontaneous nodules, in the absence of microbes (see above). This phenomenon reflects that the plant nodulation program was intact. From the above results the most plausible conclusion was that the alfalfa mutant MN-1008 had a defect in the perception of the Nod factor, that is the assumed Nod factor receptor gene suffered mutation, or one of the member in the early signal transduction chain.

30

DISCLOSURE OF INVENTION

In agreement with the patent the knowledge and use of the DNA sequence, the described technology, specifically the isolation of the Nod factor receptor gene, or individual genes in the signal transduction has a great importance in the agriculture. The transfer of these genes into plants which are unable to fix nitrogen but have great agricultural importance, plants like rice, corn, wheat, barley, rye and so on, may ensure the possibility, that in the presence of appropriate Nod factors produced by *Rhizobium* bacteria the transgenic plants carrying and expressing Nod factor receptor genes and genes of the member of the signal transduction chain might be able to perform the first biological steps leading to nodule formation. Cloning and transfer other nodule specific genes might lead to the formation of normal symbiotic nodule and nitrogen fixation.

In accordance with the above, the invention relates to the *Medicago sativa* DNA sequence encoding the NORK polypeptide as well as to the protein determined by the mentioned DNA sequence. In addition the invention features conserved DNA regions, primers, DNA probes, polypeptides and corresponding antibodies characteristic to the NORK gene family, as well as to methods and their utilization to isolate additional NORK genes from different legumes or other plants, and to use those genes for producing transgenic plants.

In another aspect, the invention features a method of cloning a DNA sequence encoding a function involved in one of the steps of symbiotic nitrogen fixation, as well as to methods of producing transgenic plant cells and cloned entire plants wherein the biosynthesis of the NORK polypeptide expresses.

The experiments described in this invention relates to the identification and isolation of genetic determinants (genes) involved in the determination and development of symbiotic nodules of alfalfa (*Medicago sativa*) and in the perception and/or signal transduction of the Nod factor originating from the bacteria, with the help of the MN-1008 mutant incapable for Nod factor triggered

nodule development, on such a way, that the mutant phenotype is genetically mapped with the use of molecular markers, followed by the isolation of overlapping BAC clones using the tightly linked markers and by recombinant DNA techniques, then sequencing the genes in the overlapping BAC clones, and finally sequencing
5 the mutant alleles responsible for the mutant phenotype. Applying the above procedure the following nucleotide sequence was identified:

```
CCATATTTTAACAATATTCTTTCTTCTACAAGGGTATAACTTTTATACAAGTTCACTATA
TTATAGGATTGATCAAGGTTCATTTTTCTTTCTTTGAAAAATCTCTAAGGGGTGTGGTT
10 TCCAAGGCAGAAAATGAAATAGAATGCAGAAGAAATTTGTATGGTACTATAAAGGGAAGAT
GAAAAGTTAGTTAGCATGGATTCAAGTTTGATAACCCTTTGGGGTAAAATCTCTTTCAGA
TTATGATGGAGCTACAAGTTATTAGGATATTTAGATTGGTTGTGGCATGTGTTCTTTGTT
TGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAGGGTTTGAGAGCATATCATGTT
GTGCTGATTCCAATTACACAGATCCAAAAACAACCCTAACTTATACAACAGATCACATCT
15 GGTCTCTGATAAAAGAAGTTGCAGACCAATACCCGAAATTTTGTTTAGCCACAGAAGCA
ATAAAAAATGTTTGAATATTTGAAATAGATGAAGGAAAGAGATGTTATACTTTGCCAACAA
TTAAGGATCAAGTATATTTGATAAGGGGTGTATTTCCCTTTGATAGTTTAAATTCTTCGT
TTTATGTTTATATCGGGGTAACAGAACTAGGTGAATTAAGATCGTCTAGACTCGAGGACT
TGGAATCGAGGGAGTTTTTAGAGCCACCAAAGACTATATTGATTTCTGCTTATTGAAGG
20 AAGATGTCAATCCCTTCATTTCTCAGATTGAATTGAGGCCATTACCTGAAGAATACCTAC
ATGGTTTCGCTACTAGTGTTTTTAAACTGATAAGCAGAAATAATCTTGGTGACACAAATG
ATGATATAAGGTTCCCAGATGACCAAAATGATAGAATCTGGAAACGGAAAGCAACTTCAA
CTCCATCATCTGCCCTTCCCCTGTCTTCCAATGTCAGCAATGTTGACCTCAAAGACAGTG
TCACACCTCCTCTACAAGTCCTACAAACAGCTCTTACTCACCTGAGCGATTGGAGTTCG
25 TCCATGATGGCCTCGAGACCGATGATTATGAATACTCTGTGTTTCTCCACTTTCTTGAAC
TAAATGGCACTGTCAGAGCAGGACAAAGGGTGTTTGACATCTATCTAAACAATGAGATTA
AAAAGGAGAAGTTTGATGTTTTGGCTGGAGGGTCCAAGAACAGTTAACTGCCTTGAACA
TTTCAGCAAATGGATCACTCAATATAACCTTAGTCAAGGCATCTGGATCTGAGTTTGGAC
CCCTTTTGAATGCCTATGAAATCCTGCAGGCACGGTCGTGGATTGAAGAGACCAACCAAA
30 AAGATTTGGAACCTTATTCAGAAGATGAGAGAAGAAGCTGCTGCTGCACAACCGAGAAAATG
AAGCATTTGGAGAGTTGGAGTGGAGACCCCTGTATGATTTTCCCCTGGAAAGGAATAACAT
GTGATGATTCAACTGGTTCATCTATTATCACTATGCTGGATCTTTCTTCCAATAATCTCA
AGGGAGCAATTCCTTACTTTGTCACTAAGATGACCAATTTACAAATACTGAACCTGAGCC
ACAACCAGTTTCGATTTCGTTATTTCCCTCGTTTCCACCGTCCTCCTTGCTGATATCATTTGG
35 ATCTGAGCTACAATGATCTTGATGGACGGCTTCCAGAATCCATTATCTCACTGCCACATT
TAAATCATTATATTTTGGCTGCAATCCATATATGAAGGACGAAGATACAACAAAGTTGA
ACAGTTCACCTAATCAATACAGATTATGGGAGATGCAAAGGAAAAAACCAAGTTTGGAC
AAGTATTCGTGATTGGAGCTATTACAAGGGGATCACTTTTGATTACTTTGGCTGTTGGAA
TTCTATTTTTTTGCCGTTATAGACACAAGTCAATTACTTTGGAAGGATTTGGTGGAAAGA
```

5 CCTACCCAATGGCAACAAATATAATCTTCTCTTTGCCAAGCAAAGACGATTTCTTCATAA
AGTCTGTATCAGTTAAACCATTTCACCTTTGGAGTATATAGAGCAGGCTACAGAACAGTACA
AAACTTTGATAGGTGAAGGAGGATTTGGTTCTGTTTACAGAGGCACTCTAGACGATGGTC
AAGAAGTGGCAGTGAAAGTGC GGTCATCCACATCAACTCAGGGAACCCNAGAATTTGATA
10 ATGAGCTAAACCTACTTTTCAGCTATACAACATGAGAACCTGGTGCCCTCTTCTGGGTTACT
GTAATGAGTATGATCAACAAATTCTCGTGTATCCATTCATGTCCAATGGCTCTTTGCTAG
ATAGACTATACGGGGAAGCATCAAAGAGAAAAATATTAGACTGGCCAACTAGACTCTCTA
TTGCTCTCGGTGCAGCTCGAGGTTTGGCATATCTTCACACATTTCCAGGACGTTCTGTAA
TACACAGGGACGTAAAATCGAGCAATATACTGCTGGATCAGAGCATGTGTGCTAAGGTTG
15 CAGATTTTGGTTTCTCAAAATACGCTCCTCAGGAAGGAGACAGTTATGTTTTCCCTTGAAG
TAAGAGGAACTGCAGGGTATCTGGATCCTGAGTACTACAAAACCCAGCAATTATCTGAAA
AAAGTGATGTTTTTCAGCTTTGGTGTGGTTCTACTTGAAATTGTAAGTGGACGGGAACCTC
TCAACATAAAGAGACCACGGATCGAGTGGAGCTTGGTTGAATGGGCTAAACCATACATAA
GAGCATCAAAGGTGGATGAAATTGTAGATCCTGGCATCAAGGGAGGATATCATGCAGAAG
20 CATTGTGGAGAGTTGTGGAAGTAGCACTGCAATGTCTAGAACCCTACTCAACATATAGGC
CATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTCATTATTGAAAACAATGCAT
CTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAACCGCTACTCAATTGTTATGG
ACAAACGGGCGCTGCCCTTCAACTACATCTACAGCAGAATCAACTATCACAACCCAAACCT
TGACACACCCCTCAACCGAGATAGTAAATGGGTCGATGGAATTCTTTTGATTTGTTTTTTA
25 TCATTGCTTTAGTAATATCCCATTTTAAATGGTAAAGGAGAAAAATACTACTTTTGATTG
TATTTTCATCCACTCTATGTTTCTTGAAACTGAATCTCTCTTGCTCAGCCCCAGTTTTTA
TGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAAACCATATGGTGCATAATTTGAAA
GCCATATTATATCATTTGCTAAGTCCAAAGTAAAAATTTACAAACTAGTTAGATTGCGA
TTTAGTCTATACACACTTCAACAGAGCTATATACACTAT

25

The invention relates to the above genomic or synthetic nucleotide sequence molecules, to the amino acid sequence deduced from this DNA, to its deletion or mutant derivatives, to its recombinant forms, and to any other proteins having NORK specific biological activity. The invention relates to those
30 oligonucleotides which are hybridizing to the above mentioned nucleotide sequence. In another aspect, the invention features those antibodies suitable for the detection of NORK polypeptides. In addition the invention relates to cells and transgenic plants as well transformed with DNA molecules protected by this invention.

35

The invention presents a procedure by which certain plants with Nod⁻ phenotype can be converted to Nod⁺ phenotype.

To achieve the aim of the invention the MN-1008 mutant plant was grown. As opposed to the wild type plant, mutant MN-1008 respond (does not start nodulation consequently does not form nodules) neither to the compatible
5 *Sinorhizobium meliloti* bacterium, nor to the Nod factor produced by the bacteria. It is supposed that the mutation in MN-1008 plant affected one of the genes involved in the signal perception and/or one of the early steps in the signal transduction pathway, it is likely the mutation affected the Nod factor receptor genes.

10 Under appropriate conditions, e.g. in the presence of sufficient nitrogen supply or in soil mutant MN1008 can grow and start to flourish. Mutant MN1008 was crossed with a nodulating wild type plant, the F1 generation and then by self-mating the F2 populations were established. In the F2 population (mapping or segregating population) the traits are segregating. From the individuals of the F2
15 population Nod⁻ and Nod⁺ individuals were collected (see Example 1.) and linked genetic markers were looked for (see Example 2.). With the help of the diploid mapping population the linked markers were mapped and it was concluded that all these markers mapped in the same region (see Example 3.). According to the map position of the diploid map, linked markers were selected from this region,
20 which were mapped in the Nod⁻ segregating population (see Example 4.). It turned out that these markers were linked to the Nod⁻ phenotype, therefore mapped in the region where the Nod⁻ mutation mapped to, and two of these were tightly linked (see Example 4.).

25 With the help of the tightly linked markers so-called BAC clones were isolated from the BAC library of *M. truncatula*, followed by the ordering of the BAC clones into so-called contigs (see Example 5.). The so-called Nod contig was about a 500 kilobase-pairs (kb) long DNA region consisting of overlapping individual BAC clones. By the analysis of about 5000 F2 plants it turned out that in
30 two individuals the recombination breakpoint situated within the Nod contig flanking both sides. This region was about 400 kb containing the Nod⁻ mutation (see Example 7.).

The BAC clones constituting the 400 kb contig were sub-cloned (see Example 6.), and the nucleotide sequence of the Nod region was determined, and the genes in this region were identified (see Example 8.). The Nod⁻ mutant phenotype could be explained by mutation in some of the genes. Without limitation only two genes are mentioned as an example. One of these genes is an ABC transporter gene. Mutation in ABC transporter genes are responsible for the multidrug resistance phenotype (34) and for the human cystic fibrosis disease (35), therefore it is possible that the Nod factor uptake is mediated by ABC transporter proteins. The other gene is coding for a receptor kinase which was named as the NORK gene. The receptor kinase genes in plants are responsible for the perception and transduction of signal molecules (36, 37, 38). The lack of one of these functions could lead to Nod⁻ phenotype. However, there are some more genes in the Nod region (see Fig. 7.) of which some are listed like the ODF gene, the MADS box gene, and the lectin genes, which could be responsible for the Nod⁻ phenotype if they carry mutation alone or in any combination.

After the sequencing of the wild type allele of these genes, the mutant alleles were also sequenced on such a way, that gene specific primer pairs were designed, synthesized and used for the amplification of the appropriate DNA region of the MN-1008 mutant plant. No nucleotide changes were found in the ABC transporter, ODF, MADS box and lectin genes which could be responsible for the mutant phenotype. Some allelic variations were found, but these mutations could be neutral changes. On the other hand, however, an in frame stop codon mutation was found in the coding region of the NORK gene originated from the MN-1008 mutant (see Fig. 9) which terminated the protein synthesis before the naturally occurring termination. As a consequence the mutant would be shorter as compared to the wild type protein, and therefore the protein is inactive. According to the sequence analysis the mutation (the stop codon) occurred in the kinase domain in a region which is functionally active and highly conserved. With the help of the RT-PCR method and specific primers (see Fig. 9.) NORK specific cDNA fragments could be amplified. According to this result NORK specific receptor

kinase mRNA was synthesized consequently the NORK gene was transcribed in both wild type and mutant plants.

It is obvious for experts in molecular biology that the mutation in the NORK gene (the *nn1* mutation which generated an in frame stop codon) leads to loss of function because the active site needed for the kinase activity was missing (see Fig. 13). It is not excluded, however, that more mutation in the MN-1008 plant occurred (like in the ABC transport gene, ODF gene, MADS box gene, and lectine gene), and the mutation leads to Nod⁻ phenotype as well, consequently, these genes are protected by this invention too.

Abbreviations used in this inventions:

NORK:	<u>N</u> OD region specific <u>R</u> eceptor <u>K</u> inase gene.
<i>nod</i> genes:	Genes found in <i>Rhizobium</i> bacteria and which are responsible for the synthesis and excretion of the Nod factor.
Nod factor:	The lipo-chitooligosaccharide produced by <i>Rhizobium</i> bacteria which induces nodule formation on plant roots (see Fig. 1.).
Nod ⁻ :	Phenotype of those plants lacking nodules
Nod ⁺ :	Phenotype of those plants with nodules
Diploid:	Somatic cells of the plants with double chromosome set
Teraploid:	Somatic cells of the plants with quadruplex chromosome set
RFLP marker:	<u>R</u> estriction <u>F</u> ragment <u>L</u> ength <u>P</u> olymorphism, which is used as genetic marker. The lenght polymorphis is detected in most cases by DNA-DNA huybridization.
SHMT:	Serine-hydroxymethyl-transferase
BAC:	<u>B</u> acterial <u>A</u> rtificial <u>C</u> hromosome
BAC vector:	Low copy number cloning vector which can accomodate large (more than 100 kb) DNA fragments.

	BAC clone:	Insert cloned into BAC vector, or both the vector and the insert together.
	BAC library:	BAC clones representing the entire DNA content of an organism.
5	bp:	base pairs.
	kb:	kilobase pairs.
	allele:	gene variant.
	primer:	short (5-50 bp long) single stranded oligonucleotide.
	contig:	DNA region covered by overlapping clones
10	in frame:	according to the open reading frame
	RT-PCR:	Reverse transcription PCR, that is amplification product of transcribed mRNA.
	NAB:	one of the F2 family segregating the Nod ⁻ character
	NBW:	the other F2 family segregating the Nod ⁻ character
15	RAPD:	<u>R</u> andom <u>A</u> mplified <u>P</u> olymorphic <u>D</u> NA (46)

BRIEF DESCRIPTION OF DRAWINGS

Fig.1. Structure of the signal molecules (Nod factors) produced by *Rhizobium* bacteria. Nod factors produced by: A. *Sinorhizobium meliloti*; B. *Rhizobium leguminosarum* bv. *viciae*; C. *Bradyrhizobium japonicum*; D. *Azorhizobium caulinodans*; E. *Rhizobium tropici*; F. *Rhizobium* sp. NGR234.

Fig. 2. F2 alfalfa populations segregating the Nod⁻ phenotype. The *nn1*/*NN1* alleles are highlighted.

25 A. Crossing of the parents; highlighting linkage group 5 with the *nn1* és *NN₁* alleles

B. The *nn1* and *NN₁* alleles inherited by the F1 NAB plant.

C. Segregating *nn1*/*NN₁* alleles in the F2 population originating from self mating of the F1 NAB plant.

30 D. Generating F3 homozygous individuals from the self mating of the appropriate F2 individuals.

Fig.3. Genetic mapping of the *nn1* gene in the tetraploid and diploid populations of *Medicago sativa*. Markers linked to the Nod⁻ phenotype were screened by the Bulk Segregant Analysis with the help of the Nod⁻ F2 NAB plants and Operon Primers. The position of the selected RAPD markers were then determined on the diploid alfalfa map. Finally RFLP markers linked to the RAPD markers were picked and mapped in the tetraploid population in relation to the *nn1* gene.

Fig. 4. Functional and restriction map of vector pBeloBac11. The cloning site (H = *HindIII*) sits in the *lacZ* gene. B = *Bam*HI; CM = chloramphenicol, see ref (52).

Fig.5. BAC clones in the Nod contig. Primary clones isolated by the help of the Q5E and SHMT markers are highlighted by green. Blue colors represent secondary clones isolated with the help of the end sequences of the primary clones. „b” and „j” represent the left and right side of the inserts, respectively.

Fig. 6. Determination of the precise position of the *nn1* gene with the help of two close recombination events. The precise localization of the *nn1* gene was established with the help of two DNAs originating from individuals carrying recombination events in the Nod region, as well as with the help of two DNAs originating from opposite homozygous individuals. Sub-clone G34P44 originating from BAC 67A11 was used as hybridization probe for filter containing EcoRV digested DNA from NAB4156 (recombinant), NAB814 (homozygous), NAB4443 (recombinant), and NAB2161 (homozygous) individuals. Nod⁺ allele (highlighted by + arrow) could be detected for the two Nod⁺ plants (NAB4156 and NAB 814), while Nod⁻ allele for the two Nod⁻ plants (NAB4443 and NAB2161). Sub-clone G33P135 has been used to identify a Nod⁻ allele for NAB4156. Sub-clone G3P126 identified Nod⁻ allele for plant NAB4443 after *Dra*I digestion. Sub-clone G18P4 on the other hand detected a Nod⁺ allele for this plant after *HindIII* digestion. Since the two plants displaying recombination were Nod⁻ the *nn1* gene can be limited between these two recombination points.

Fig.7. Genes and their orientation found in the Nod region. The arrows show the orientation of the genes in the Nod region. The lengths of the arrows are out of scale. pNORK_Nhe5 is a pUC19 based recombinant plasmid carrying the

entire NORK gene on a 8.5 kb long NheI fragment. The name of the genes are given after the abbreviations. The accession number of that gene is given which showed the highest BLST score in the NCBI databank. Finally the e value is given, which is a generally accepted value representing the level of homology.

5 Fig. 8. Nucleotide sequence of the 8563 bp long DNA region containing the entire NORK gene. The transcription start site (ATG) is at position 1153 bp (left to right) from the NheI site (GCTAGC). The stop codon is at position 663 bp from (right to left) the last nucleotide of the second NheI site. There are 15 exons and 14 introns in the gene. The position of the exons and introns are highlighted in
10 Table 6. (There are 60 nucleotides in one row).

Fig. 9. List of the oligonucleotide primers used for the amplification of both genomic and cDNA sequences.

A. Name, sequence, length, and start points of the primers.

15 B. Schematic representation of the NORK gene with the position of the primers.

Fig. 10. cDNA sequence of the NORK gene from *Medicago truncatula* A17. (There are 60 nucleotides in one row).

20 Fig.11. Alignment of the genomic (see Fig 8.) and cDNA (see Fig. 9.) sequence of the NORK gene from *Medicago truncatula* A17. The genomic sequence (BAC A17) of the NORK gene identified in *Medicago truncatula* A17 was aligned to the cDNA sequence (cDNS A17). For the genomic sequences corresponding to the BAC clones: intron sequences are highlighted by red, the rest are black (Promoter region, exons, 5' and 3' untranslated regions). cDNA sequences are highlighted by blue. Nucleotide sequences corresponding to the first two (GT) and last two (AG) nucleotides in the introns are double underlined. Start codon (ATG) is green, stop codon (TAG) is purple. NheI sites at the beginning and at the end of the sequences are highlighted by orange.

30 Fig. 12. Deduced amino acid sequence from the NORK cDNA of *M. truncatula* A17 plant. (* = stop codon).

Fig.13. Functional representation of the NORK protein by highlighting the characteristic domains.

- A. The NORK protein is 925 aa long. The different structural/functional domains are highlighted by different colors.
- B. More detailed description of the functional domains found in the NORK protein.

Fig.14. cDNA sequence of the NORK gene for the two Nod⁻ alleles 1 and 2, in plants (NAB1241/6 and NAB701/28) of *Medicago sativa*

Fig. 15. cDNA sequence of the NORK gene for one of the Nod⁺ alleles (allele 6) in plant (NAB615/28) of *Medicago sativa*.

Fig. 16. Deduced amino acid sequence of the cDNA for the NORK gene from Nod⁻ plants (NAB1241/6 és NAB701/28) of *M. sativa*. (* = stop codon).

Fig.17. Deduced amino acid sequence of the cDNA for the NORK gene from Nod⁺ plants (NAB615/28) of *M. sativa*. (* = stop codon).

Fig. 18. Alignment of the cDNA sequences of the NORK genes originating from *Medicago truncatula*, *Medicago sativa*, *Vicia villosa* and *Pisum sativum*. *Medicago* alleles are from *Medicago truncatula* A17 (MtA17), *M. sativa* Nod⁻ allele (Ms1N⁻) and Nod⁺ allele (Ms6N⁺), as well as from *Vicia villosa* (Vv) and *Pisum sativum* (Ps). *M. sativa* Nod⁻ alleles 1 and 2 were identical. Coding regions are grouped in three nucleotide groups according to the genetic code. The position of the introns are highlighted by light purple (i1-i14). Nucleotides are highlighted by red if they are different from the MtA17 sequence. Mutations creating stop codons are shown by arrows. At least two alleles are in *V. villosa*. Ambiguous nucleotides are highlighted as: M=A or C; R=A or G; W=A or T; S=C or G; Y=C or T and K=G or T.

Fig. 19. Alignment of the deduced amino acid sequence of the NORK gene from *M. truncatula* A17, *M. sativa* allele 6 and allele 1 as well as from *Vicia villosa* and *Pisum sativum*. *Medicago* alleles are from *Medicago truncatula* A17 (MtA17), *M. sativa* Nod⁻ allele (Ms1N⁻) and Nod⁺ allele (Ms6N⁺), as well as from *Vicia villosa* (Vv) and *Pisum sativum* (Ps). *M. sativa* Nod⁻ alleles 1 and 2 were identical. Amino acid differed from the MtA17 sequence are highlighted by red. At least two alleles are in *V. villosa*. The corresponding amino acids are separated by /. Stop codons are highlighted by red. (There are 60 aa in one row).

Fig. 20. RFLP pattern of the NORK gene for the individuals from the diploid mapping population to demonstrate the single copy nature of the NORK gene. Hybridization pattern of the of the parents (*M. sativa* ssp. *coerulea* w2 és *M. sativa* ssp. *quasifalcata* k93), the F1/1 and individuals of the F2 generation on a filter containing DraI digested total DNA and hybridized by NORK specific probe (see Example 10.). The DNA used for probe contains two DraI sites, which are probably present in the NORK gene of the diploid species. As a consequence there are at least three hybridizing bands. (Other DraI sites are not excluded in the intron sequences of the genomic DNA). The 212 bp long DraI fragment probably did not transfer to the filter. The 10.5 kb fragments were not polymorphic, as contrast to the small one which is polymorphic. According to the segregation of this fragment there is only one copy of the NORK gene in the diploid alfalfa. Mcw2=*Medicago sativa* ssp. *coerulea* w2; MqK93=*Medicago sativa* ssp. *quasifalcata* k93

Fig. 21. RFLP hybridization of the NORK gene from tetraploid alfalfa. Hybridization pattern of the Nod⁻/Nod⁺ *M. sativa* individuals from the segregating population. Total DNA was digested with HindIII, the probe contained only the extracellular part of the cDNA. On the autoradiogram, there is only one strong hybridization band for four individuals, and there are two bands for one individual. Plant NAB814 carries only alleles 5, and 6, respectively. The rest of the plants carry only allele 1 and 2 between the recombination sites (see Example 7.). Allele 5 and 6 can be distinguished according to the size of the two hybridization

fragment in NAB814. On the other hand, allele 1 and 2 are indistinguishable (only one hybridizing band). This hybridization pattern indicates that there is only one copy of the NORK gene in the tetraploid alfalfa as well.

5 Fig.22. Hybridization pattern of the *M. truncatula* total DNA to the NORK gene. Hybridization pattern of *M. truncatula* total DNA digested by EcoRI and EcoRV enzymes. The probe ("NORK specific") contained only the extracellular part of the cDNA. Hybridization bands represent sequences homologous to the NORK gene.

10

Fig.23. Hybridization pattern of total DNA from different legumes (*Sesbania*, *Cassia*, *Trifolium* és *Desmodium*) to the NORK gene. The hybridization pattern is shown on filters containing total DNA digested by *EcoRI*, *EcoRV*, *DraI* and *HindIII* enzymes. The probe ("NORK specific") contained only the extracellular part of the cDNA (see Example 10.). Hybridization bands represent sequences homologous to the NORK gene. S. r. = *Sesbania rostrata*; C. e. = *Cassia emerginata*; T. p. = *Trifolium pratense*; D. sp. = *Desmodium* species.

Fig 24. Hybridization pattern of total DNA from different legumes (*Vicia*, *Trifolium* és *Melilotus*) to the NORK gene. The hybridization pattern is shown on filters containing total DNA digested by *EcoRI*, *EcoRV*, *DraI* and *HindIII* enzymes. The probe ("NORK specific") contained only the extracellular part of the cDNA (see Example 10.). Hybridization bands represent sequences homologous to the NORK gene. V. s. = *Vicia sativa*; M. a. = *Melilotus alba*; T. p. = *Trifolium pratense*; T. i. = *Trifolium incarnatum*.

Fig 25. Hybridization pattern of total DNA from different legumes (*Vigna* és *Macroptilium*) to the NORK gene. The hybridization pattern is shown on filters containing total DNA digested by *EcoRI*, *EcoRV*, *DraI* and *HindIII* enzymes. The probe ("NORK specific") contained only the extracellular part of the cDNA (see Example 10.). Hybridization bands represent sequences homologous to the NORK

gene. V. u. = *Vigna unguiculata*; M. a. = *Macroptilium atropurpureum*; V. r. = *Vigna radiata*

Fig 26. Hybridization pattern of total DNA from different legumes (*Pisum*,
5 *Glycine*, *Lotus*) to the NORK gene. The hybridization pattern is shown on filters containing total DNA digested by *EcoRI* enzyme. The probe ("NORK specific") contained only the extracellular part of the cDNA (see Example 10.). Hybridization bands represent sequences homologous to the NORK gene.

10 Fig 27. Hybridization pattern of total DNA from different non-legumes (rice, tobacco, wheat and corn) to the NORK gene. The hybridization pattern is shown on filters containing total DNA digested by *EcoRI* enzyme. The probe ("NORK specific") contained only the extracellular part of the cDNA (see Example 10.). Hybridization bands represent sequences homologous to the NORK gene.

15 Fig.28. cDNA sequence of the NORK gene from *Pisum sativum* cv. Frisson plant. (There are 60 nucleotides in one row).

Fig. 29. cDNA sequence of the NORK gene from *Vicia villosa* S-1 plant.
20 (There are 60 nucleotides in one row). The nucleotides represented by the two allele of *V. villosa* are highlighted as follows: M = A or C, R = A or G, W = A or T, S = C or G, Y = C or T, K = G or T.

Fig.30. Deduced amino acid sequence of the cDNA for the NORK gene
25 from *P. sativum* cv. Frisson plant. Stop codon (*) is highlighted by red. (There are 60 aa in one row).

Fig. 31. Deduced amino acid sequence of the cDNA for the NORK gene
30 from *V. villosa* S-1 plant. At least two alleles are in *V. villosa*. The corresponding amino acids are separated by /. Stop codon (*) is highlighted by red. (There are 60 aa in one row).

Fig. 32. Antigen-antibody reaction using diluted (1:1000) rabbit serum with cell protein extracts from 10 mg plant tissue. A. control tobacco plant (non-transformed). B. Transgenic plant carrying the MtA17 NORK gene. C. Nod⁺ control alfalfa NAB814. D. Nod⁻ alfalfa (MN-1008).

5

Fig. 33. Multiple alignment of the NSL proteins. Amino acid residues were coloured according to an 80 % consensus. Letters mark the following amino acid classes: h = hydrophobic (ACFGHIKLMRTVWZ) and l = their aliphatic subset (ILV) with turquoise shading; a = aromatic (FWY) with green shading; + = positive (HKR) with red shading; p = polar (CNEHKNQRST) in red; t = "turn-like" (ACDEGHKNQRST) in green; s = small (ACDGNPSTV) in blue. White letters with dark violet background mark amino acids identical (conserved 100%) in all sequences. Amino acids present in 80% of the sequences are shown by white letter with pink background. Stars indicate the positions where the amino acids of the legume NORKE domain differ significantly from the consensus of other NSL sequences. Mt = *Medicago truncatula*, At = *Arabidopsis thaliana*, Pv = *Phaseolus vulgaris*, Os = *Oryza sativa*, Gm = *Glycine max*, Gos = *Gossypium*, Zm = *Zea mays*, TC = tentative consensus sequences constructed from EST sequences by TIGR (www.tigr.org).

Table 1. Germination ability, viability and nodulation properties of the F2 seeds for the NAB and NBW families. F2 seeds were germinated on 1 % water-agar, after germination seedlings were transferred to tubes containing slant agar (Gibson). After one week plants were inoculated with *S. rhizobium* 41. After 6 weeks and 2 months nodulation was evaluated visually. For detailed description see Example 1.

Table 2. Genotyping the F2 NAB and NBW individuals for markers OPW8a, OPE8c, OPB13b, OPA6a and for nodulation. Total DNA from the highlighted NAB (15 Nod⁻ ; 15 Nod⁺) and NBW (8 Nod⁻; 22 Nod⁺) individuals were amplified by OPW8, OPE8, OPB13, OPA6 Operon Primers as described in Example 2. During

the amplification reaction fragments (RAPD fragments) corresponding to markers OPW8a, OPE8c, OPB13b and OPA6a either appear (coded by number 5) or not (coded by number 3). Genotypes of the Nod⁺ and Nod⁻ individuals are coded by number 5 or 3, respectively. Genotypes in the table are highlighted according to the concept of colormapping, colors without number are predicted genotypes (see ref. 39).

Table 3. Genotyping the OPW8a, OPE8c, OPB13b, OPA6a markers on the diploid mapping population. Genotyping of the maternal and paternal alleles corresponding to the hybridizing bands were done as follows: maternal homozygous = 1 (yellow); paternal homozygous = 3 (purple); heterozygous = 2 (green); maternal dominant = 5 (light green); paternal dominant = 4 (blue); missing genotypes = 0 (grey). Colormapping was according to Table 3, prediction according to ref. 39. Colormaps show the U584B-LbIII region of linkage group 5 of the diploid *M. sativa* ssp. *coerulea* and *M. sativa* ssp. *quasifalcata* genetic map (40). 3A and 3B highlights the non-predicted and the predicted map, respectively.

Table 4. Genetic mapping of markers U71, U224, CG13, SHMT, Q5E és U492 on the tetraploid alfalfa population. Genotypes of the individuals plants were according to the size of the hybridizing band. If the band was similar to the size of the band from the Nod⁻ parent, the genotype code was 3 (purple). If the band was similar to the size of the band from the Nod⁺ parent, the genotype code was 5 (light green). Genotypes of the Nod⁺ and Nod⁻ individuals are coded by number 5 or 3, respectively. Genotypes in the table are highlighted according to the concept of colormapping, colors without number are predicted genotypes (see ref. 39).

Table 5. List of the sub-clones originating from BAC clones 28I12, 50E23, 2D11 és 67A11. The size of the insert in the sub-clones are also shown. For detailed description see Example 6. kb = kilobase pairs.

Table 6. Position of the exons and introns in the NORK gene according to Fig.9. Numbering starts from the first nucleotide of the NheI site (see Fig. 8.).

Table 7. Cloning of the NORK gene into plant cloning vectors. The description of the vectors used for plant transformation and the details of the cloning is given in Example 12.

5

Table 8. List of the *Agrobacterium* strains carrying the NORK gene. The construction of the *Agrobacterium rhizogenes* and *A. tumefaciens* strains used for the two different transformation protocols are given in Examples 12A and 12B, respectively.

10

Table 9. List of the *A. rhizogenes* strains used for transformation and the transformation frequency. Nod⁻ plants were infected by *A. rhizogenes* strains as described in Example 12B. After emergence of the roots GUS staining was performed and the nodulation ability was tested. The second and the third row in the Table highlights the results obtained for the two transformation methods (combined). In brackets, the results from the second experiment are shown.

15

Table 10. Summary of the transformation experiments with *Medicago sativa* and *Agrobacterium tumefaciens*.

20

Table 11. Summary of the transformation experiments with *Nicotiana tabacum*, *Vicia villosa* and *Agrobacterium tumefaciens*.

BEST MODE OF CARRYING OUT THE INVENTION

25

Hereinafter the invention is demonstrated by examples. These examples do not restrict the protected domain of the invention they are only for demonstration purposes.

Example 1

30

Genetic crosses and the analysis of F2 progeny of the NAB and NBW families

The tetraploid ($2n=4x=32$) *Medicago sativa* MN-1008 mutant line (28) has been provided us by Prof. D. Barnes (University of Minnesota, MN, USA) (primarily called MnNC-1008 (NN), but generally abbreviated as MN-1008), is
5 unable to form symbiotic nodules in the presence of compatible bacteria. The mutant plant emerged spontaneously by means of crosses between cultivated alfalfa populations and subsequent self-matings (27). One individual (referred in laboratory notebook as MN-1008/17) of the purple flowered MN-1008 mutant
10 alfalfa line has been used as maternal parent in genetic crosses. The paternal parent was one individual of the *M. sativa* cv. Nagyszénási (referred as MsNa/5) alfalfa population. Since the self-pollinating efficiency of the MN-1008/17 maternal parent was between 10-20 %, its flowers had to be sterilized (emasculated). Prior to crosses the so-called banner petals of the matured flowers of the maternal
15 parents were trimmed by scissors and followed by splitting open of the flowers by squeezing them to avoid the anthers to hit the stigma. Then the flowers were immersed in 51 % ethanol for 3-4 sec and then washed in water. After the flowers dried the stigma was covered with the pollens from the male parent. The efficiency of the ethanol treatment was checked in control experiments: the self-pollinating efficiency of the treated flowers in the absence of foreign pollen was about 0.5 %.

20 The F1 progeny seeds coming from the above described cross were vernalized at 5-8 °C for two weeks, then germinated and planted out first into plastic pots with diameter 8 cm, and after the plants gained strength into pots with diameter 25 cm containing flower-soil termed NT. Plants are maintained to bloom in greenhouse at 22-25 °C with photoperiod of 16 hours lighting and 8 hours dark.
25 The flowers were self-pollinated one by one manually by squeezing the base of the flowers (tripping). At this time the stamen knocked against the stigma and the pollens got into the stigma from the anthers. After self-pollination the F2 ovules appeared on the plants from which the matured seeds developed. The F2 generation were begun to nurse in symbiotic test to determine their symbiotic
30 phenotype. The matured seeds were vernalized and their surface were sterilized by rinsing them in 70% ethanol for 30 sec and treating with 0.1% HgCl₂ solution for 5 min followed with rinsing in sterile distilled water for five times. Then the

seeds were swelled in sterile water for 8 hours. After swelling, the seeds were placed in equable distances from each other on the surface of plates in Petri dishes containing 1% agar in distilled water. The dishes were incubated upside down in dark to prevent the emerging roots to grow into the agar medium. Next day the seedlings with roots of about 0.5 – 1.0 cm length were then placed on the surface of 1% agar slants with nitrogen-free Gibson medium (41) and the plants were grown at 22 - 25 °C with 16-h photoperiod (10 000 Lux). The seedlings were inoculated with *Sinorhizobium meliloti* 41 bacteria culture suspended in 1ml Gibson medium after 5-7 days of transfer.

10 The *S. meliloti* 41 bacteria culture were prepared for inoculate as follows. An inoculation loop of *S. meliloti* 41 strain (42, 43) stored at -80 °C were inoculated into test tube containing 5 ml YTB medium (44). The bacteria culture was rotated in test tube for 24 hours. The grown-up culture were inoculated into a 500 ml flask containing 100 ml YTB medium and the culture were grown up to stationer phase for about 24 hours. The concentration of the bacteria suspension grown-up to stationer phase was between $1-3 \times 10^9$ cells/ml. The bacteria were collected by centrifugation (in centrifuge Sorvall RC5 with GSA type angle rotor at 4 °C with 10 krpm for 10 min) under sterile conditions. The supernatant was discarded and the cells were resuspended in 100 ml Gibson medium (41). The concentrated bacteria culture was diluted in 100-fold in Gibson medium before inoculation. After dilution the concentration of the bacteria was $1-3 \times 10^7$ cells/ml. The nodulation phenotype of the plants was scored 6 weeks after inoculation, afterwards plants were potted into soil. Plants were removed gently out of the pots and were evaluated once more for the presence or absence of nodules 2 months later (secondary control of nodulation). The plants were planted back into pots and grown further. The flowers of the F2 individuals selected later by genetic analysis (determination of the genotypes for different molecular markers) were self-pollinated as described above to generate F3 seeds which were tested for symbiotic nodulation.

30 After the crosses 62 F1 seedlings were planted and grown up. Forty-four of these plants produced sufficient flowers and F2 seeds (at least 50-100 seeds/individuals) after self-pollination to test their viability and analyze the

segregation of symbiotic nodulation phenotype. Two F1 plants (NAB and NBW) producing the most vigorous F2 progeny in which the Nod⁻ phenotype could be evaluated were selected for further experiments. The germinating, viability and nodule forming characters of NAB and NBW F2 seeds are listed in Table 1. The
5 segregation ratio of the Nod⁻ and Nod⁺ characters was between 1:50 and 1:55 that is comparable to the segregation ratio of a single, recessive character in a tetraploid F2 segregation population (theoretical value is 1:35).

Composition of plates containing 1 % agar in distilled water:

10 10 g Bacto agar

1 liter distilled water

Sterilization for 30 min at 121 °C

Composition of flower-soil termed NT:

15 1 part Florina flower-soil (commercially available)

1 part peat (commercially available)

1 part sand from river Maros (commercially available)

Composition of YTB medium:

20 1% Bacto Tryptone

0.1% Yeast extract

0.5% NaCl

1 mM MgSO₄

1 mM CaCl₂

25

Example 2

Identification of linked RAPD markers to the *nn₁* gene of the Nod⁻ MN-1008 mutant plant using the NAB and NBW F2 families

30

The identification of tightly linked markers to the *nn₁* gene, that is to the Nod⁻ character was performed with the so-called Bulk Segregant Analysis method

(45). To perform the screen five Nod⁻ and five Nod⁺ plants were selected from the NAB and NBW families, respectively (Nod⁻ NBW plants: NBW6, NBW20, NBW37, NBW162, NBW790; Nod⁺ NBW plants: NBW9, NBW23, NBW74, NBW440, NBW718; Nod⁻ NAB plants: NAB65, NAB267, NAB637, NAB701, NAB908; Nod⁺ NAB plants: NAB26, NAB82, NAB139, NAB421, NAB814). The analysis of two F2 families enhanced the possibility to identify more linked markers. Since alfalfa is an outcrossing plant, the genome is heterozygous carrying different alleles for the genes. Analysing the inheritance of the alleles in a tetraploid population (as shown in Figure 2.) demonstrated that NAB F1 individual contained two Nod⁻ alleles (indicated as 1 and 2) and two Nod⁺ alleles (indicated as 5 and 6), while it did not carry the two other Nod⁺ alleles (indicated as 7 and 8). As it turned out later, the analysis of the NBW family allowed us to detect polymorphism of these Nod⁺ alleles (indicated as 7 and 8) and to identify additional linked molecular markers.

Total DNA was isolated from young leaves of the Nod⁻ and Nod⁺ plants using the QIAGEN Plant DNA Isolation Kit (DNeasy Plant Mini Kit; QIAGEN Inc. – USA; 28159 Avenue Stanford; Valencia CA 91355) according to the supplier's instructions. The concentration of the DNA samples were determined by spectrophotometer (OD₂₆₀= optical density measured at 260 nm) (1 OD₂₆₀ = 50 µg/ml DNS). The DNA samples were diluted to a concentration of 5 ng/µl, and the DNA samples of five Nod⁻ and Nod⁺ plants were mixed, to produce the Nod⁻ and Nod⁺ homozygous groups. The mixed DNA samples of the homozygous groups were used as templates in polymerase chain reactions (PCR): 12.8 µl sterile distilled water, 5 µl DNS, 2.5 µl 25 mM MgCl₂, 2.5 µl 10 x Taq polymerase buffer (Zenon Kft, H-6720. Szeged, Berzsenyi u. 3.), 2 µl Operon Primer (2.5 pmol/µl), and 0.2 µl Taq DNS polymerase (5 U/µl) (Zenon) were pipetted into PCR tubes of 500 µl volume. The PCRs were carried out in 40 cycles with the following steps: 5 sec at 94 °C, 1 min at 37 °C and 1 min at 72 °C. After completion of the PCR amplification, 5 µl AGE (AGE = 20 % sucrose, 0.1 % bromophenol blue, 0.1 % xylene cyanol) stain was added to the samples and then the amplified products were separated in 2 % agarose gel (Sigma Low EEO) of 15 x 20 x 0.4 cm with 2 V/cm voltage for three hours in 1x TEA (40 mM Tris, pH 7.8, 50 mM EDTA, 50 mM acetic acid) buffer. After separation of the fragments, the gel were stained in 0.1 %

ethidium bromide solution and exposed to UV light to take pictures. The first 520 10-mer primers of Operon Primer set (Operon Technologies, Alameda, Ca, USA) with different nucleotide composition were tested to amplify of the DNA of the two homozygous bulks for both the NAB and NBW families. The PCR amplifications
5 with these short primers resulted in fragment pattern typical for the given primer. The detected differences in the pattern (polymorphism) could be used as RAPD molecular markers (46). As a result of the tests of the primers of the Operon Primer set, several amplified fragments could be identified that were amplified on the bulk of the Nod^+ DNA but not on the bulk of the Nod^- DNA. Four out of them
10 were analysed thoroughly: OPA6a (identified in NBW family), OPW8a (identified in NAB family), OPE8c és OPB13b (identified in both families). The sizes of the amplified products identified in the bulks of the Nod^+ plants were as follows: OPA6a, about 650 bp; OPW8a, about 300 bp; OPB13b, about 850 bp; OPE8c, about 480 bp. Afterwards the the individuals DNAs were re-tested in PCR to
15 identify the genotypes. In addition other individuals from the families (15 NAB and 8 NBW Nod^- plants, as well as 15 NAB and 22 NBW Nod^+ plants) were tested for the same four RAPD markers. The result of the genotyping (thegenotypes) was displayed in color-map (39) in Table 2. The results presented in the table demonstrated unambiguously the linkage between the nn_1 gene and the four
20 RAPD markers and determined the genetic order of the markers.

For fine mapping, the NAB family has been selected since it was the largest F2 population. The F2 individuals with diallelic configuration have been selected from this family to produce F3 individuals by self-mating (see Figure 2.). Individuals with homozygous genotypes for the different alleles were searched by
25 determination of their genotypes. Three F3 plants were selected: the plant designated as NAB1241/6 that is Nod^- and contains the alleles labelled with no. 1 in homozygous configuration, the Nod^- plant designated as NAB701/28 containing the alleles labelled with no. 2 in homozygous configuration around the genomic region of the mutation and the Nod^+ plant designated as NAB615/28 carrying the
30 alleles labelled with no. 6 in homozygous configuration. No individuals could be detected containing homozygous alleles labelled with no. 5 in this region.

Example 3

Genetic mapping of OPW8a, OPE8c, OPB13b és az OPA6a RAPD markers on the linkage map of diploid alfalfa (*M. sativa*)

5

In order to determine the map position of the *nn₁* gene on the linkage map of alfalfa, the closely linked RAPD markers (markers OPW8a, OPE8c, OPB13b and OPA6a) were mapped on the genetic map of diploid alfalfa (40). To execute this experiment, the genotypes of the individuals of the diploid segregation population had to be determined. DNA fragments were amplified as described above using the OPW8, OPE8, OPB13 and OPA6 primers, respectively and the total DNA of NAB814 and NBW9 plants as template DNA. After separation, the DNA fragments corresponding to markers OPW8a, OPE8c, OPB13b and OPA6a were isolated from the agarose gels, purified by the QIAGEN Fragment Isolation Kit (QIAEX II Gel Extraction Kit; Cat.No. 200 51 QIAGEN Inc) and their concentration determined by spectrophotometer (determination of OD at 260 nm). Fifty ng of the fragments were labelled with deoxycytidine-5'-[alfa-³²P] triphosphate (Code: FP-205, Izotóp Intézet Kft., 1121 Budapest, Konkoly Thege Miklós út 29-33.) using the Ready-To-Go DNA labelling (-dCTP) Kit (cat. number: 27-9240-01; Amersham Pharmacia Biotech Inc, Piscataway, NJ, USA) according to the suppliers' instructions. The HybondN⁺ nylon filters (cat. number: RPN3003B, Amersham Pharmacia Biotech Inc, Little Chalfont, Buckinghamshire, England) carrying the restriction enzyme digested DNA of the individuals of the diploid mapping population were hybridized with the hybridization probes prepared as described above. The filters were prepared by the following method: total DNA was isolated of the individuals of the F2 mapping population (47) as described in Example 2. Aliquots of 20 µg DNA were digested with *Dra*I, *Eco*RI, *Eco*RV and *Hind*III restriction enzymes (Fermentas AB, Vilnius, Lithuania), respectively and the restriction fragments were separated in 0,8 %-os agarose gels as described in Example 2. The fragments were transferred onto HybondN⁺ nylon membranes according to the suppliers' instructions and the filters were hybridized at 60 °C for 16 hours in CG buffer published by Church and Gilbert (48). After hybridization,

30

washing was carried out three times for 20 minutes each at the temperature of the hybridization in the following solutions: 1.) 0.1% SDS and 1xSSPE (49); 2.) 0.1% SDS and 0.5xSSPE; 3.) 0.1% SDS and 0.1xSSPE. Autoradiography of the filters was at -80°C using the Sterling Diagnostic (Sterling Diagnostic Imaging Inc.,
5 Newark, DE 19714 USA) X-ray films or the filters were exposed in the exposure cassette of the phosphorimager scanning instrument (PhosphorimagerTM 445 SI; Molecular Dynamics Inc., Sunnyvale, CA, USA) and after the scanning of the screens the results were analysed by the ImageQuant computer software. The hybridizing fragments corresponding to the alleles of the OPW8a, OPE8c, OPB13b
10 and OPA6a markers were identified as RFLP bands and the genotypes of the individuals were determined. The map position of the markers were determined by the colormapping procedure (39): all of the four markers were mapped in linkage group 5 (LG5) close to the U224 marker. The genotypes of the OPW8a, OPE8c, OPB13b and OPA6a markers of the individuals in the diploid mapping population
15 and the results of the genetic mapping are presented in Table 3.

Example 4Genetic mapping of the RFLP markers between U584B and LbIII markers in the tetraploid population segregating the *nn₁* gene

5

To ensure that the map position identified in the diploid population with molecular markers linked in fact to the *nn₁* gene (to the Nod⁻ phenotype) the map position of some markers mapped previously in this region (between the U584B and LbIII markers) were determined in the tetraploid alfalfa genetic map. The segregation of five RFLP markers (U71, U224, CG13, SHMT and U492) and a specific PCR marker (marked with Q5E) generated by cloning the amplified fragment (preparing the pQ5E plasmid) of the RAPD marker designated as OPQ5E were analyzed in the NAB population segregating the *nn₁* gene:

Preparation of pQ5E plasmid: PCR amplification was performed using 50 ng total DNA of NAB814 plant (see the preparation of plant total DNA in Example 2) and OPQ5 primer as described in Example 2. The amplified products were separated in agarose gel (see Example 2) and the 1750 bp-long fragment corresponding to the Q5E marker were excised, reisolated from agarose gel and purified as described in Example 3. After adding 5.5 μ l 10x Klenow buffer to the 50 μ l eluted DNA solution, fragments were treated with 0.5 μ l (5 unit) Klenow fragment of the DNA polymerase I according to the suppliers' (Fermentas AB) instructions. Then 4 μ l (1 μ g) *HincII*-digested [prepared according to suppliers' (Fermentas AB) instructions], CIP (Calf Intestine Alkaline Phosphatase)-treated (dephosphorylation) and purified pUC19 plasmid (49), 7 μ l 10x ligase buffer and 3 μ l (15 U) T4 DNA Ligase enzyme (Fermentas AB) were added to the fragments. The ligation was performed at 14 °C for 24 hours. After this the ligation was transformed into 250 μ l *E. coli* DH5 α competent cells (Amersham Pharmacia Biotech AB, Uppsala, Sweden) using the CaCl₂ procedure (50). After induction of the bacterial gene which resulted in resistance against ampicillin (phenotypical expression), the bacteria were plated on LB plates containing ampicillin (100 μ g/ml) and incubated at 37 °C for 16 hours. Plasmid DNA was isolated from the resistant colonies by QIAGEN plasmid isolation kit (QIAGEN Plasmid Mini Kit; Cat.

No. 12123 QIAGEN Inc.). One clone was selected by PCR amplification from the clones carrying the 1750 bp-long amplified product: BRC1464 = *E. coli* DH5 α (pQ5E). The sequence of the insert was determined and specific primers were designed based on the sequence:

- 5 Q5EU1: 5'-CGGCGTCTTGCTAAAGGAGA-3';
- Q5ED1: 5'-TATTCGACTCATCATGGTTA-3';
- Q5EU2: 5'-GTCAAATATCGATTCTGTGAT-3';
- Q5EU3: 5'-CTAAGAAAGGCTTTGGTTGG-3';
- Q5ED3: 5'-TCACGAATCGATATTTG ACA-3').

10 Using the specific primers in PCR amplification, the dominant OPQ5E marker was transformed into codominant Q5E marker and its more precise map position was determined in relation to the other markers.

The map position of markers U71, U224, CG13, Q5E és U492 have been published earlier (40). The SHMT marker (51) mapped between markers Q5E and

15 U224 on the genetic map of diploid *M. sativa* (Kaló et al. manuscript in preparation). Markers U71, U224, CG13, SHMT and U492 were mapped as RFLP markers using the F2 NAB individuals, while the marker Q5E was mapped by specific PCR (94°C 30sec, 60 °C 1 min, 72 °C 1 min in 40 cycles) as well as by

20 RFLP in the tetraploid population. The inserts of the clones (47, 51) corresponding to RFLP markers and pQ5E clone were reisolated, labelled, hybridized, washed and autoradiographed according to the method described in Example 3. The filters were selected for hybridization containing the *Dra*I, *Eco*RI, *Eco*RV and *Hind*III-digested total DNA of the selected Nod⁻ és Nod⁺ NAB individuals. The genotypes of the individuals were determined for markers U71, U224, CG13, SHMT, Q5E and

25 U492 by the hybridization signals corresponding to the different alleles. The order of the genes shown in Table 4 was determined by the colomap procedure. The summary of the genetic mapping of the *nn₁* gene displaying the steps from the identification of the linked markers to the confirmation of the map position in the tetraploid population is presented on Figure 3. Analyzing the mapping data, it

30 turned out that the markers SHMT and Q5E were the most tightly linked markers to the *nn₁* gene. No recombination could be detected in the individuals of the

analyzed population between markers SHMT and nn_1 . One recombination could be detected between markers nn_1 and Q5E in the analyzed individuals.

Example 5

Identification of overlapping BAC clones, constructing the contig

5 The primary BAC clones (67A11, 2D11 és 20K10) were isolated by two molecular markers (SHMT and Q5E, see Figure 3.) most tightly linked to *nn₁* mutation in the following way: the Q5E fragment was isolated from plasmid pQ5E, while the SHMT fragment was excised from the plasmid containing the cDNA of the gene with *KpnI* restriction enzyme (51). The fragments were reisolated after gel
10 electrophoresis, labelled and the filters were hybridized, washed and autoradiographed as described in Example 3. The filters contained more than 30 000 clones of the BAC library constructed from the total DNA of *Medicago truncatula* a close relative of *Medicago sativa*, (53). The BAC clones in the library contained about 100 kb inserts of the total DNA of *M. truncatula*. Cloning was
15 done by *HindIII* digestion of the total DNA and cloned into the *HindIII* restriction sites of the pBeloBACII vector (52) (Figure 4.). The filters and clones of the BAC library are available in a public collection (Clemson University Genome Institute, 100 Jordan Hall, Clemson 29634-5727, South Carolina, USA).

Two primary BAC clones (clones 67A11 and 2D11) was identified by
20 hybridization with SHMT fragment, and one primary BAC clone (20K10) with the Q5E fragment. An inoculation loop of the BAC clones 67A11, 2D11 and 20K10 stored in the clone collection at -80 °C were taken from the stock and inoculated into 500 ml flask containing 100 ml LB media and chloramphenicol in 15 µg/ml concentration, respectively. The clones were grown to stationer phase at 37 °C in a
25 shaker. (250 rpm). The plasmid DNA of BAC clones were purified by the QIAGEN Large Construct Kit according to the suppliers' instructions. The plasmid DNA of the three clones were subjected to single *EcoRV* and double *EcoRV/NotI* digestion and the fragments were separated in agarose gels. Based on the pattern of the single and double digestion, the *EcoRV* end-fragments of the clones could be
30 selected. From these fragments the end fragment of the inserts could be detected by *HindIII* digestion. These end-fragments were used for chromosomal walking, that is to identify overlapping BAC clones. The *HindIII* end-fragments of the clone

67A11 were approximately 3300 és 800 bp, of the clone 2D11 were 3300 and 4000 bp and of the clone 20K10 were appr. 1080 és 1400 bp length, respectively. The six end-fragments were reisolated from agarose gel and repeated hybridization was performed to identify and isolate the secondary BAC clones (50E23, 64B5, 28I12, 6A5). Based on the restriction digestion pattern (*HindIII*, *EcoRI*, *EcoRV*) of the primary and secondary BAC clones, the degree of the overlap of the clones could be determined and the position of the clones to each other could be established. In this way the Nod contig could be constructed (see Figure 5.).

10

Example 6

Subcloning of the Nod contig

15 The result of the genetic mapping revealed that the SHMT marker could be the closest marker to the *nn₁* gene therefore the chromosomal region around the SHMT was analyzed first thoroughly by delimiting the genomic region containing the *nn₁* gene and by determination of the sequence of this region. In parallel experiments the subcloning of two BAC clones (67A11, 2D11) identified by SHMT and the subcloning of clones 50E23 és 28I12 overlapping with 2D11 clone were
20 needed. The subcloning was performed in three ways. On one hand (A) the ends of the BAC inserts in the contig were cloned, on the other hand (B) the 1-2 kb length fragments of the clones 67A11, 2D11 and 50E23 generated by sonicating were cloned, so random subcloning was performed, and on the third hand (C)
25 cloning was realized after restriction enzyme digestion.

Example 6A

Subcloning of the end-fragments of the BAC inserts

30

The end-fragments of the BAC clones were identified as described in Example 5 and the *HindIII* fragments were cloned into pUC19 plasmid in the

following way: 4 μ l (1 μ g), *Hind*III-digested and dephosphorylated pUC19 plasmid (Fermentas AB), 14 μ l 5x Ligase buffer, and 2 μ l (10 unit) T4 DNA Ligase enzyme (Fermentas AB) were added to the DNA fragments purified with the QIAGEN DNA Isolation Kit (see Example 3) and eluated in 50 μ l buffer. The ligation was executed
5 at 14 °C for 24 hours. The ligation was transformed into 250 μ l *E. coli* DH5 α competent cells using the CaCl₂ procedure (see Example 4). The transformant colonies were checked by the restriction enzyme digestion of the plasmid DNA isolated with QIAGEN plasmid DNA Isolation Kit (see Example 3).

10 Example 6B

Random subcloning of the inserts of the 67A11, 50E23 and 2D11 BAC clones

15 20 μ g DNA of the clones 67A11, 50E23 and 2D11 in 200 μ l TE buffer (1 mM EDTA, 10 mM Tris-HCl, pH7.5), respectively, were sonicated for 10 sec with the smallest head of the MSE Ultrasonic Disintegrator equipment. The fragmented DNA was separated in agarose gel and the 1-2 kb length fragments were excised, reisolated and purified as described in Example 3. The DNA fragments generated
20 in this way were ligated into *Hinc*II-digested (Fermentas AB) and dephosphorilated pUC19 plasmid as described in Example 4 and then they were transformed into *E. coli* DH5 α competent cells (see Example 6A). Plasmid DNA were isolated from the transformant colonies with QIAGEN Plasmid Isolation Kit and their contents for inserts were checked by electrophoresis in 1 % agarose gels.

25

Example 6B

Subcloning of the 67A11, 2D11, 50E23 and 28I12 BAC clones after restriction enzyme digestion

30

The plasmid DNA of BAC clones 67A11, 2D11, 50E23 and 28I12 were purified as described in Example 5 and 1 μ g purified plasmid DNA were digested

with *NheI*, *XbaI*, *Hin2I*, *Hin6I*, *TaiI*, *TaqI*, *BamHI*, *HindIII*, *EcoRI* and *SauIIA1* enzymes (Fermentas AB), respectively according to the suppliers' recommendation. The *NheI* and *XbaI* fragments were cloned into *XbaI*-digested (the *NheI* and *XbaI* result in compatible sticky ends) and dephosphorylated pUC19
5 vector, the *Hin2I*, *Hin6I*, *TaiI*, *TaqI* fragments were cloned into *AccI*-digested and dephosphorylated pUC19 vector; the *HindIII* and *EcoRI* fragments were cloned into *HindIII*- and *EcoRI*-digested and dephosphorylated pUC19 vector and the *BamHI* and *SauIIA1* fragments were cloned into *BamHI*-digested and dephosphorylated pUC19 vector. After digestion the restriction fragments were purified by the
10 QIAGEN DNA Isolation Kit (see Example 3), ligated into the corresponding enzyme-digested (*NheI*, *AccI*, *HindIII*, *EcoRI* and *BamHI*) and dephosphorylated pUC19 plasmid and transformed as described in the Example 6A. The transformant colonies were checked by digestion with the corresponding restriction enzyme after plasmid DNA isolation as described above (Example 6A).
15 The recombinant clones and their insert sizes generated with subcloning are presented in Table 5.

Example 7

20 Determination of the map position of the *nn₁* gene more precisely in the contig by identification of the closest recombination sites at both sides of the mutation

The generation of the subclones from the BAC clones overlapping the *nn₁*
25 mutation and the determination of the map position of the *nn₁* gene more precisely by this subclones were done parallelly. The generated subclones were used as genetic markers to determine the recombination sites more precisely. After having the recombination sites more precisely it was obvious that the detailed sequence analysis for BAC 67A11 and 2D11 was necessary.

30 Two Nod⁻ plants (NAB4156 és NAB4443) carrying recombination close to the *nn₁* mutation and two other plants that did not contain recombination close to the Nod region but they were opposite homozygotes (Nod⁺ NAB814 and Nod⁻

NAB2161) for the Nod character were used for genetic experiments. The clones G34P44 and G33P155 coming from BAC clone 67A11, the clone G3P126 coming from 2D11 and clone G18P4 arising from 28I12 BAC clone were converted to molecular markers. The inserts of the clones were excised with the proper
5 restriction enzyme, the fragments were reisolated and used as hybridization probes to the filters containing the *EcoRV*-, *HindIII*- and *DraI*-digested total DNA of above mentioned four plants. From the experimental results (see Figure 6) it could be concluded that the position of the *nn₁* gene was delimited on a genomic region (Nod region) by two recombination detected by fragment G34P44 located on BAC
10 clone 67A11 and by fragment G18P4 located on BAC clone 28I12.

Example 8

Determination and analysis of the sequence of the Nod region

15

In order to determine the sequences of the subclones obtained as described in Example 6, the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit and the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer Applied Biosystems; 850 Lincoln Centre Drive Foster City, CA94404
20 USA) was used according to the manufacturer's recommendation. The plasmid DNA for templates of the sequencing reaction were isolated by QIAGEN Plasmid Isolation Kit and their concentration were quantified using spectrophotometry (see Example 3.). The sequences of the amplified products labelled by fluorescent dyes were determined by ABI 373 and ABI 377 automated sequencer (Perkin
25 Elmer Applied Biosystems; 850 Lincoln Centre Drive Foster City, CA94404 USA). Authenticity of the DNA sequences: the sequences were determined on both strand, in one direction the reactions were repeated 2-8 times, based on independent templates in ambiguous cases. The sequence data were stored in computer and their analyses were started with ordering them into overlapping
30 sequences. The overlapping end-fragments of the BAC clones and of the subclones generated by restriction digestion of these BAC clones could help to assemble the sequences and to arrange the subclones generated by random

fragmentation or restriction enzyme digestion and the BAC clones to each other. By the help of the available sequences, homologue genes were searched in the NCBI (National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/BLAST/>) and the *Arabidopsis* (<http://www.arabidopsis.org>) databanks. Taking into consideration of the homology between the sequences, the common and general characters of the gene structures [consensus sequences as initiation and termination codons, open reading frame (ORF), consensus sequences characteristics of exons and introns such as GT-AG rule, the point of divergence, etc.] the starting- and end-point of the coding regions of the genes located in the Nod region were searched and the order and the orientation of the genes were determined. The results of this sequence analysis are presented schematically on Figure 7.

Example 9.

Detailed analysis of the NORK gene

The Nod⁻ phenotype of the mutant plant can be explained above all with a mutation in the NORK (NOD region linked Receptor Kinase) gene coding for a receptor kinase, that is why the sequence analysis of the wild type NORK gene on BAC clone 67A11 was carried out first. The NORK gene is located on an NheI fragment of 8563 bp length and its nucleotide sequence - from the NheI site to the next NheI site - is shown in Figure 8. This nucleotide sequence identify the genomic sequence of the NORK gene from the *M. truncatula* A17 plant (the BAC library was constructed from this plant). To reveal the exact structure of the gene we have to know the exon/intron borders/positions which can be predicted from the genomic sequence based on the GT-AG rule and the branching point, however, it can be determined exactly only by knowing the cDNA sequence. For this purpose we isolated RNA from the roots of *M. truncatula* line A17, then we prepared cDNA from the mRNA by reverse transcription. cDNA samples were produced by PCR

amplification using NORK specific primers(see Figure 9.), then the amplified fragments were sequenced either directly or after cloning as described in Example 8.

5 Example 9A.

Construction and sequencing of NORK specific cDNA sequences

The seeds of *M. truncatula* A17 were germinated after sterilization and
10 planted into sterile soil. From the roots of the two-three week old plants the cDNA
representing the NORK mRNAs as well, was prepared as follows. The roots of the
plants were pulverized/ground under liquid nitrogen, the powder was solubilized in
RNA extraction buffer (8 M guanidinium chloride, 20 mM MES, 20mM EDTA, 50
mM β -mercaptoethanol), then purified by extraction with a mixture of phenol-
15 chlorophorme-isoamil-alcohol (25:24:1). After centrifugation the supernatant was
precipitated at -20 °C for 1 hour with 0.7 volume of 95% ethanol and 0.2 volume of
1 M acetic acid. After centrifugation the precipitated RNA was dissolved in DEPC
(diethyl-pyrocabonate)-treated distilled water. The cDNA was prepared from the
RNA with the help of the RevertAid H Minus First Strand cDNA Synthesis Kit
20 (Fermentas AB). The different RNS samples were treated with RNase free DNase
(Fermentas AB) and then the cDNAs were synthesized with the help of oligo-dT
primer and reverse transcriptase enzyme (Fermentas AB) in the presence of
RNase inhibitor (Fermentas AB) at 42 °C. Overlapping fragments of the NORK
cDNA were amplified from the prepared cDNA with different combinations of the
25 primers shown in Figure 9. The PCR conditions were as follows: 95 °C for 30
seconds, 55 °C for 1 minute, 72 °C for 1 minute, these steps were repeated 35-
times. The specific NORK fragments amplified by RT-PCR from the cDNA were
separated by agarose gel electrophoresis and purified with the help of the
Fragment Isolation Kit of QIAGEN. The purified fragments were sequenced as

described in Example 8. either directly or after cloning them into HincII digested and phosphatase treated pUC19 vector. The full-length cDNA sequence could be constructed from the sequences of the overlapping cDNA fragments. The cDNA sequence of the NORK gene from the *M. truncatula* A17 plant is shown in Figure 5 10.

Example 9B.

The alignment and analysis of the genomic and cDNA sequences of the
10 NORK gene.

The exact position of the exons and introns could be determined by aligning the genomic and the cDNA sequences of the NORK gene. The alignment of the NORK gene sequences from the *M. truncatula* A17 plant is shown in Figure 11., the details about the position of exons and introns are presented in Table 6. 15 The proposed start codone of the NORK gene is the ATG start codone (coloured green) at the 1153. bp of the *NheI* fragment (pNORK_*NheI*5), while the proposed stop codone is the TAG stop codone (coloured violet) at the 7898. bp. In the putative promoter region at the 5' end of the gene can be found the TATA-box (743-55, 769-85) és CCAAT-box (675-79, 688-91 bp) sequences characteristic for the region. At the 3' end of the gene there are several sequences which may serve as transcriptional termination signals. There is no difference between the two sequences (genomic and cDNA) which is in accordance with the facts that the genomic and the cDNA is originated from the same plant, and there is only one 20 gene in the genome.

By comparing the sequences to the entries of databases we realized that in the case of *M. truncatula*, in the so-called EST (Expressed Sequence Tags) databases there are partial cDNA sequences homologous to the NORK gene. These sequences were obtained by random sequencing, i.e. certain laboratories

made sequencing reactions and single runs on several thousand cDNA clones of different libraries without any knowledge about their function. Such sequences from the *M. truncatula* plant can be found under accession numbers AW684661, AW685681 and BE203249. The appearance of the cDNA sequences in this form
5 also support the expression of the NORK gene in the roots of legumes.

The activity of the NORK gene in the cells, i. e. it is expressed, is proven by the fact that we could detect NORK specific cDNA corresponding to the mRNA via RT-PCR. We presented as described in Example 14. that the NORK protein is synthesized from the mRNA. The amino acid sequence of the NORK protein can
10 be deduced from the cDNA sequence. The amino acid sequence deduced from the longest Open Reading Frame (ORF) of the NORK cDNA is shown in Figure 12. The NORK protein consists of 924 amino acids, and it contains characteristic sequence motifs pointing to functions, which were revealed on the basis of similarity to consensus sequences in databases. The functional representation of
15 the NORK protein and its characteristic sequence motifs are shown in Figure 13.

From the above data obtained in this way we concluded that one part of the molecule is located extracellularly (extracellular part) which is followed by an intracellular part via a transmembrane stretch. According to our knowledge the extracellular part contains a so-called LRR (Leucin Rich Repeats) region
20 (according to the homologies: between amino acids 405-476) which may be involved in either the direct or the indirect recognition and in either the direct or the indirect binding of Nod factor. Based on the similarities, the proposed intracellular part has kinase activity. It is known for a specialist of molecular biology that the kinases participate in different signal transduction pathways, they initiate or
25 transduce the biological signals by phosphorylating proteins. The NORK protein may phosphorylate its substrates within the cell and thus it initiates the signal transduction cascade leading the development of the nodule. Based on the structural analysis of the protein it is possible to suppose that the NORK protein is capable for autophosphorylation.

Example 9C.

The alignment of the cDNA sequence of *M. truncatula* A17 to those of the
5 mutant (Nod⁻) and the wild type (Nod⁺) *M. sativa* plants.

With the help of primers designed to the NORK gene of *M. truncatula* (Figure 9.) genomic and cDNA sequences (Example 9A.) were amplified from F3 plants carrying in homozygous configuration the Nod⁻ alleles designated as number 10 1 and number 2 (NAB1241/6 and NAB701/28), as well as the Nod⁺ allele designated as number 6 (NAB615/28). The nucleotide sequences of the amplified products were determined directly or after cloning as described in Examples 8. and 9A. The cDNA sequence of the Nod⁻ alleles designated as number 1 and number 2, as well as the Nod⁺ allele designated as number 6 is shown in Figures 14. and 15., respectively, the amino acid sequences deduced from the cDNA sequences of the Nod⁻ alleles designated as number 1 and number 2, as well as the Nod⁺ allele designated as number 6 can be seen in Figures 16. and 17., respectively. The sequence analysis revealed that the sequences of the NORK alleles designated as number 1 and number 2 are identical, that is why only one sequence is presented.

20 The alignment of the cDNA sequences of the Nod⁻ and Nod⁺ alleles from *M. sativa* to the cDNA sequence of the *M. truncatula* NORK gene is shown in Figure 18. One can see from the comparison of the sequences that there are a number of differences between the sequences of *M. sativa* and that of *M. truncatula*, and a few differences - in most cases causing no amino acid changes - can be found 25 between the wild type and mutant alleles of *M. sativa*. The most dramatic change in the structure of the Nod⁻ allele carrying the *nn₁* mutation can be found at the 11. basepair in the 13. exon (E13) of the gene. The mutation in the triplet coding for tyrosine resulted in a STOP codone in the sequence originated from the Nod⁻ plant. As a result of this mutation, the synthesis of the protein is terminated earlier, even

before the synthesis of the putative active site of the kinase, and it can be supposed that the produced shorter protein cannot fulfil its function in the signal transduction..

The alignment of the amino acid sequences of the NORK proteins of the *M. truncatula* A17 és a *M. sativa* plants is shown in Figure 19. The comparison of the sequences reveals that the sequences of *M. sativa* differ at a number of positions from that of *M. truncatula* A17. The amino acid differences between the Nod⁺ alleles of *M. truncatula* A17 and *M. sativa* are probably the results of so-called neutral mutations which do not change the function. The comparison of the amino acid sequences of the Nod⁻ and the Nod⁺ NORK alleles reveals that the most dramatic change in the structure of the NORK protein in the *M. sativa* mutant MN-1008 is that the protein consists of only 709 amino acids because there is a STOP codon in place of the tyrosine codon in position 710 which causes the termination of protein synthesis (see above). There are more than one amino acid changes in the protein originated from the Nod⁻ allele compared to that of the Nod⁺ allele which might result in the loss of function. It cannot be decided that in which order the mutations in the *nn₁* gene arose and which one was the first that in homozygous configuration resulted in the fail of nodule formation, i.e. in Nod⁻ phenotype.

20 Example 10.

The number of NORK genes in the *Medicago* genome and its distribution in the plant kingdom

25 The genetic mapping is a suitable tool to determine the copy number of the NORK gene. For this purpose, using the putative extracellular part of the NORK gene (this part is specific, there is no homologous sequence in the databasis) as a hybridization probe, hybridization experiments were carried out as described in Example 3. with the filters carrying the DNA of the individuals from the diploid and

tetraploid mapping populations of *M. sativa*. The 1695 bp length probe, which is called as „NORK specific” probe, was prepared by PCR using primers Pr_RKU4 és Pr_RKD2X (Figure 9.) and a cDNA template originated from the *M. truncatula* A17 plant as described in Example 9A. The analysis of the autoradiograms showed that the segregating hybridizing fragments identify a single locus (Figures 20. and 21.), which means that there is only one copy of the NORK gene in the genome, i.e. the NORK gene is a single-copy gene.

The above described probe was used for the hybridization to the genomic DNA of different plants. With the help of the QIAGEN Plant DNA Isolation Kit genomic DNA was isolated as described in Example 2. from different legume plants: *M. truncatula* A17 és A20, *Pisum sativum* cv. Frisson, *Sesbania rostrata*, *Cassia emerginata*, *Desmodium* sp., *Vicia sativa*, *Melilotus alba*, *Trifolium pratense*, *Trifolium incarnatum*, *Vigna unguiculata*, *Macroptilium atropurpureum*, *Vigna radiata*, *Glycine max*, *Lotus corniculatus*, *Lotus japonicus* cv. Gifu, and from non-legume ones: *Nicotiana tabacum* cv. Small Havana SR1, and rice. 15-20 µg of total DNA was digested by restriction enzymes *EcoRI*, *EcoRV*, *DraI* or *HindIII*, the DNA fragments were separated by electrophoresis and bound to Hybond N+ filters (see. Example 3). The autoradiograms obtained from the hybridizations performed at 55 °C using the „NORK specific” probe (Example 10.) can be seen in Figures 22., 23., 24., 25., 26., 27. One can see on the autoradiograms that there is only one or two strongly hybridizing fragments in most legumes which means that in most of the legumes the NORK gene is a low copy number gene, only one or few genes are present in the genome. The autoradiograms obtained after the hybridization experiments with non-legume plants show that in these species the presence of the NORK gene cannot be detected under our hybridization conditions.

Example 11.

Determination of the NORK cDNA sequence from other legumes.

The nucleotide sequences of the NORK gene were determined from a single individual of two other legume species, pea (*Pisum sativum*) and vetch (*Vicia villosa*). The sequencing of the NORK cDNA from *Pisum sativum* cv. Frisson and *Vicia villosa* S-1 plants was carried out using the primers designed for the NORK gene of *Medicago truncatula* (Figure 9.) as described in Example 9A. Both *Pisum sativum* cv. Frisson and *Vicia villosa* S-1 has Nod⁺ phenotype, similarly to *Medicago truncatula*, diploid, autogamous plants, which means that the two alleles of their NORK gene has identical sequence. The nucleotide sequence of the amplified products was determined directly or after cloning as described in Examples 8. and 9A. Figures 28. and 29. show the NORK cDNA sequences of *Pisum sativum* cv. Frisson and *Vicia villosa* S-1, respectively, constructed from the sequences of overlapping fragments. Figures 30. and 31. show the amino acid sequences deduced from the cDNA sequences of these plants.

The cDNA sequences determined by us from *Pisum sativum* and *Vicia villosa* plants as well as the cDNA sequence from the *Lotus japonicus* Miyakojima MG-20 plant deposited in the NCBI database under the Accession number AV410167 proves that the NORK gene is transcribed in these plants as well, i.e. it is active. It is noteworthy to mention that in *Vicia villosa* there are either two alleles or two copies of the NORK gene (see Figures 29. and 31.). It can be concluded as well that the nucleotide sequence of the NORK gene in plants investigated by us is so homologous to that of *M. truncatula* that the primers designed for the NORK gene of *M. truncatula* could be used for the amplification of different parts of the NORK gene from alfalfa, pea and vetch.

Example 12.

Complementation of the Nod⁻ mutant *M. sativa* MN-1008 plant with the NORK gene.

In order to prove that the identified mutation in the NORK gene is responsible for the Nod⁻ mutant phenotype, the 8.5 kb *Nhe*I fragment carrying NORK gene (Table 5., Figures 7. and 8.) was cloned into plant transformation vectors (see Table 7.). The NORK gene was cloned in the so-called T-DNA part of the *Agrobacterium*-based plant transformation vectors. *Agrobacterium* species that entered the tissues of plants transfer the T-DNA from their plasmid into the plant cells (T-DNA transfer), which is followed by the integration of T-DNA (54). The random integration of T-DNA into different regions of the plant genome makes possible the stable maintenance and inheritance, as well as the expression of the genes of the T-DNA. There are two basic methods to introduce T-DNA based plasmids into plants: the *Agrobacterium tumefaciens* and the *Agrobacterium rhizogenes* mediated transfer. Using *A. tumefaciens* strains, transformed calli are obtained first, from which somatic embryos have to be induced which have to be regenerated to raise transgenic plants. This procedure can be carried out only by using so-called embryogenic plants. Using *A. rhizogenes* strains for infection, hairy roots grow on the infected part of the plant, and as a result a chimeric organism is formed in which only the hairy roots are transformed.

Since preliminary results showed that the plant (MN-1008) carrying the *nn₁* mutation is not embryogenic under standard conditions, i.e. plants cannot be regenerated from the calli transformed by the *A. tumefaciens* mediated method, two strategies were chosen for the genetic complementation.

(i) Using the *Agrobacterium rhizogenes* transformation method, when the ability for the symbiotic nodule formation of the transformed hairy roots can be investigated. Transformed roots of most legume species are able to form nodules(55, 56), that is why, in the case of complementation, the formation of nodules might be expected on the roots of the mutant plants.

(ii) The non-embryogenic MN-1008 plant was crossed as described in the Example 1. with an individual of an embryogenic line (Regen SY) which was provided by Dr. Deborah A. Samac, University of Minnesota, MN, USA (57). After the selection and self-pollination of the embryogenic F1 individuals, the symbiotic phenotype of the individuals of the F2 population was determined as described in the Example 1., and two plants from the Nod⁻ progeny (F2RN28/4, F2RN28/5) as well as two embryogenic F1 plants (F1RN28 és F1RN41) was used for the *A. tumefaciens* mediated transformation experiments.

10 Example 12A.

The cloning of the NORK gene into plant transformation vectors

The fragment carrying the NORK gene was cloned into several plant transformation vectors which have different copy number and provide different DNA-environment in the T-DNA that might affect the expression of the transformed gene.

The 8.5 kb NheI fragment carrying the NORK gene was cloned into the different plant transformation vectors from plasmid pBRC1660 which was constructed as follows: From the G20P5 clone (see Table 5.) constructed as described in Example 6., the insert was cut by the restriction endonucleases *PaeI* and *SacI* and was ligated into the pOK12 (49) vector which was digested with the same enzymes and dephosphorilated by the CIP enzyme. From the obtained clone designated as pAT688 the insert was cleaved by the restriction endonucleases *ClaI* és *SacI* and it was ligated into the vector pBlueScript II SK (Stratagene, 11011 North Torrey Pines Road, La Jolla, CA, USA). From the clone pBRC1660 obtained as described above (maintained in *E. coli* DH5 α strain under the code BRC1660) the NORK gene can be deliberated by several enzymes. The pOK12 vector was provided by Dr. J. Messing (Waksman Institute, Piscataway, NJ, USA).

The 11.8 kb *ClaI* fragment harbouring the NORK gene with a 2.2 kb longer 5' (promoter) sequence was cloned into the different plant transformation vectors from plasmid pBRC1690 which was constructed as follows: The BAC clone 67A11 shown in Figure 5. was digested by *ClaI* enzyme, then the fragment was re-
5 isolated from agarose gel as described in the Example 3. and was ligated into pBlueScript II SK vector which was digested by the same restriction endonuclease and dephosphorilated by the CIP enzyme.

The pPK459 vector which was used for the construction of the pBRC1678 clone was developed as follows: The *HindIII-EcoRI* fragment from the polylinker of
10 the pSL301 plasmid (59) was cloned between the *HindIII* és *EcoRI* sites of the pGA471 vector described by An et al. (58). The pGA471 vector (58) and the pSL301 vector (59) were kindly provided by G. An (Washington State University, Pullman, WA, USA) and J. Brosius (Mt. Sinai School of Medicine, New York, NY, USA), respectively.

15 The pPR97 vector (60) used for the construction of clones pBRC1666 and pBRC1701 was provided by László Szabados (MTA, BRC, Szeged).

The pAT680 vector used for the construction of the pBRC1667 clone was constructed as follows: the T-DNA part of the pGREEN II vector described by Hellens et al. (61) was digested by *HpaI* enzyme and treated by the CIP enzyme
20 and was ligated to an *EcoRV* fragment carrying the nopaline synthase gene promoter (nos) driven neomycin phosphotransferase II gene providing kanamycin resistance (nos-KmR cassette), and the obtained clone was deposited in the strain collection under the name pAT678. After removing by *EcoRV* digestion from the carrier plasmid, the intron-containing *uidA* gene (62) coding for the β -
25 glucuronodase gene and driven by 35S promoter of the cauliflower mosaic virus (35S-GUS-INT cassette; 61) was built into the *StuI* digested CIP treated pAT678 vector. The T-DNA containing *BglII* fragment was cloned from the obtained pAT679 plasmid into the *BamHI* site of the broad host range vector pAT672 which was constructed from the replication origo of plasmid pBBR1MCS-4 (63) and the

kanamycin resistance gene of plasmid pK19 (64) as follows: The kanamycin resistance gene from plasmid pK19 was amplified as described in Example 4. using primers KMU (5'-GGCGATCGATAGACTGGGCGG) and KMD (5'-TCGTGATGGCAGGTTGGGCG), then the amplified fragment was cleaved by *Clal* and reisolated from agarose gel as described in Example 3. The isolated fragment was cloned into the *EheI* and *Clal* digested pBBR1MCS-4 vector. The original restriction enzyme cutting sites from the resulted pAT671 vector were removed in two steps as follows: The DNA was digested by *BamHI* and *Bsp119I*, then *BamHI* and *Clal* enzymes and the digested DNAs were treated in the presence of 0,05 mM dNTP with the Klenow fragment of DNA polymerase I as described in Example 4. and ligated. The overhanging ends of the DNA resulted after the *BamHI* and *Bsp119I*, as well as the *BamHI* and *Clal* digestions were filled in by the Klenow fragment which restored the recognition site (GGATCC) of *BamHI*. Plasmids pK19 (64) and pBBR1MCS-4 (63) were kindly provided by D. Pridmore (Ciba-Geigy AG, Basel, Switzerland) and E. Kovach (Louisiana State University, Shreveport, LA, USA), respectively. Vector pGREEN II and cassettes nos-KmR and 35S-GUS-INT can be obtained from the John Innes Centre, Norwich, England (detailed information: www.pgreen.ac.uk).

These cloning vectors (pPK459, pAT680, pPR97) were digested by restriction enzymes shown in the third column of Table 7., purified and dephosphorilated by the CIP enzyme as described in Example 4. After a second purification step 1 µg of vector DNA was ligated with 1 µg of the insert carrying the NORK gene. The insert carrying the NORK gene was deliberated from the plasmid pBRC1660 by enzymes shown in the third column of Table 7. then isolated from gel as described in Example 3. The ligated DNAs were transformed into *E. coli* strain DH5α, then after the checking of the transformant colonies (see Example 6A) the *E. coli* strains carrying the clones were deposited under the codes BRC1666, BRC1667, BRC1678 and BRC1701 (Table 7.).

The plasmids carrying the NORK gene (pBRC1666, pBRC1667, pBRC1678, pBRC1701) were purified from the *E. coli* strains and were transformed into the different *Agrobacterium rhizogenes* és *Agrobacterium tumefaciens* strains shown in Table 8. The transformations were carried out by electroporation as follows.

5 Bacteria grown on solid LB medium containing 100 µg/ml of rifampicin antibiotics were inoculated into 5 ml of YENB (65) medium containing 100 µg/ml of rifampicin and were incubated overnight at 30 °C with good aeration. 200 ml of YENB medium was inoculated with this starter culture. The bacterium culture was grown with intensive aeration to late logarithmic phase (optical density at 600 nm,

10 OD₆₀₀=0.6-0.8), then the cells were collected by centrifugation. The bacteria were washed at 0°C two times with 100-100 ml of sterile distilled water and two times with 10 ml of 1mM HEPES (pH=7) containing 10% of glicerol. Bacterial cells were resuspended at 0°C in 2-4 ml of sterile 1mM HEPES (pH=7) containing 10% of glicerol, and these electrocompetent cells were frozen in liquid nitrogen as 100 µl

15 aliquots and stored at -80 °C till usage. 0.1-1 µg of plasmid DNA was added to the electrocompetent cells on ice, then the DNA was introduced into the cells at 1.25 kV/cm field strength with the help of to E.coli Pulser instrument (BioRad) according to the supplier's instructions. Immediately after the electroporation, 1 ml of LB medium was added to the cells, then after an incubation at 30 °C for 120 minutes

20 the cells were plated onto LB medium containing 100 µg/ml of rifampicin and antibiotics shown in Table 7.

Example 12B.

25 Transformation using *Agrobacterium rhizogenes* strains

The *A. rhizogenes* strains (BRC1673, BRC1675, BRC1679; Table 7.) carrying the NORK gene on plasmid can be introduced into the tissues of the Nod

mutant *M. sativa* plants by two ways: 1.) through the sectioned surface of seedlings' radicle and 2.) through the sectioned surface of shoots of mature plants.

Transformation via cutting the radicles of seedlings: The seeds of the MN-1008 and the Nod⁻ plants (NAB701, NAB809, NAB897, NAB968, NAB2792) carrying *nn*₁ mutation - which were identified and maintained as described in Example 1, - were sterilized and germinated on 1% water agar as described in Example 1. The rootlets of the 3-4 day old seedling were cut 5-10 mm below cotyledon and the wounded roots were dipped into a lawn of *A. rhizogenes* grown on solid LB medium. The *A. rhizogenes* strains (BRC1673, BRC1675, BRC1679) were streaked from the strain collection stored at -80 °C onto LB medium containing 100 µg/ml of rifampicin and antibiotics shown in Table 7. and were incubated at 30 °C for 3 days. The plants treated with the *Agrobacterium* strains were placed into squared (10x10 cm) Petri-dishes containing TM-1 medium (66) and were grown at 24°C with 16 hour light period and 10000 Lux light intensity. After 5 days roots were appearing at the cut surface, and part of them were hairy root, i.e. transformed root.

2. Transformation through the sectioned shoots of mature plants: The young shoots of the MN-1008 and the Nod⁻ plants (NAB701, NAB809, NAB897, NAB968, NAB2792) identified and maintained as described in Example 1. that carry the *nn*₁ mutation in homozygous configuration were cut from the plants and placed immediately into tap water. The sectioned surfaces of the shoots were dipped into melted paraffine to avoid the entering of the chemicals used for sterilization into the vascular tissues. Then the shoots were sterilized as follows: they were washed in 70% ethanol for 5 seconds then immediately in 20% bleach containing 0.05 % Tween 20 for 90 seconds. After 3 washing in sterile destilled water the shoots were cut into pieces of 3-5 cm. After dipping the root-proximal surface into *Agrobacterium rhizogenes* lawn grown as described in the previous paragraph the distal end of the plants were stabbed into TM-1 medium and were incubated at 22°C with 16 hour light period and 10000 Lux light intensity. From the 20. day

roots were appearing at the place of the wounding and some of them were transformed

Transformed roots could be identified by the so-called GUS-staining (67) because the T-DNA carries the uidA gene coding for the β -glucuronidase enzyme.

5 The roots were stained for the GUS activity as follows: root pieces were incubated at 37°C in dark for 12-24 hours in a buffer containing 100 mM TRIS (pH=7.4), 50 mM NaCl, 2 mM spermidine, 2 mM K₃Fe(CN)₆, and 2 mM X-GlcU (5-bromo-4-chloro-3-indol β -D-glucuronic acid) dye. The β -glucuronidase enzyme expressed in the transformed roots cleaves the dye and as a result a blue product precipitated in

10 the cells is formed which can be detected both by eyes and by microscope. The *A. rhizogenes* strains used for transformation and the transformation efficiency is shown in Table 9. After the transformation 4 transformed roots were obtained, however, no nodule formed in the presence of compatible *S. meliloti*. The reason for this results might be that the T-DNA of the Ri plasmid harbours genes coding

15 for enzymes involved in plant hormone biosynthesis and the expression and biological effect of these genes in the plant cells inhibit nodule formation (Á. Kondorosi, personal communication).

Example 12C.

20

Transformation using *Agrobacterium tumefaciens* strains

The *A. tumefaciens* transformation was carried out as follows: Twenty pieces of healthy, dark-green leaves were cut from the alfalfa plant to be

25 transformed and placed immediately into Petri dishes containing tap-water. The leaves were sterilized as follows: after washing with 70 % ethanol for 5 seconds, they were placed immediately into a 20 % of bleach containing 0.05 % of Tween 20, and washed for 90 seconds. This was followed by three washing steps in sterile distilled water. The edges and the central vascular tissues of the sterilized

leaves were removed and the two pieces were split in two. These leaf pieces - with all the edges cut - were placed into SHO medium containing *A. tumefaciens* which was prepared as follows: The *A. tumefaciens* strains (BRC1677, BRC1680, BRC1681, BRC1707; Table 8.) stored at -80 °C were streaked on solid medium containing 100 µg/ml of rifampicin, as well as the proper antibiotics shown in Figure 7. and incubated at 30°C. A single colony was inoculated by loop into 3 ml of liquid YEP medium (in 20 ml tube) and the bacteria were grown by rolling (to provide aeration) to the stationary phase (24 óra). The bacteria were collected by centrifugation as described earlier, the supernatant was poured away, and the bacteria were suspended in 12 ml of SHO medium. After keeping in the *Agrobacterium* suspension for 20 minutes, the leaves were removed and the excess of suspension was blotted by sterile blotting paper, then the leaf pieces were placed onto B5h in Petri dishes (maximum 16 leaf pieces per plate). The plates were sealed by air-permeable bands and placed to 24°C and 16 hour light period with 10000 Lux light intensity. After 1 week the leaves were removed from the medium, washed 3 times with sterile distilled water and placed onto B5hKmCb medium then were incubated as described above. After 3 weeks the leaf pieces forming calli were placed onto B5hKmCb medium and were further incubated as described above. The appearing dark green embryos were transferred onto MMSCb medium and were further incubated as described above. The plantlets forming roots and shoots were transferred into jars containing MMSCb medium where the putative transformants were raised. The developed plants were propagated in vitro as follows: a piece of the shoot containing 3-4 pair of leaves was cut under sterile conditions, placed into MMSCb medium in a jar and incubated as described above.

After strengthening of the shoots and roots of the propagated plants, DNA was isolated from the leaves of the plants as described in Example 2., then to detect the transformation events PCR reactions were carried out as described in the Example 3. by using primers specific for the *nptII* gene providing resistance

against kanamycin (NPT-U: 5'-ACCCAGCCGGCCACAGTCG-3', NPT-D: 5'-GGGCGCCCGGTTCTTTTGTG-3') and for the uidA gene encoding the β -glucuronidase enzyme (GUS-U: 5'-TTATGCGGGCAACGTCTGGTAT-3', GUS-D: 5'-AGTCCCGCTAGTGCCTTGTCC-3') that are within the T-DNA.

- 5 After transferring to nitrogen-free Gibson medium, the plants were infected with *S. meliloti* strain 41 and the process of nodulation was followed. Transformed derivatives of RN28 and RN41 were self-pollinated then germinated as described in Example 1. and we checked their nodule formation ability.

 The results of the transformation experiments by using the *A. tumefaciens* strains are summarized in Table 10.

10

The composition of the YEP medium:

10 g/l Protease Peptone (Difco Laboratories)

10 g/l Bacto yeast extract (Difco Laboratories)

5 g/l NaCl

5 Sterilization: 121 C°, 20 minutes.

The composition of the LB medium:

10 g/l Bacto-tryptone (Difco Laboratories)

5 g/l Bacto yeast extract (Difco Laboratories)

10 5 g/l NaCl

pH=7.5

Sterilization: 121 C°, 20 minutes.

The composition of the YENB medium:

15 8 g/l Bact Nutrient Broth (Difco Laboratories)

7.5 g/l Bacto yeast extract (Difco Laboratories)

Sterilization: 121 C°, 20 minutes.

The composition of the SHO medium:

20 Schenk and Hildebrandt salt:

KNO₃ 2500 mg/l

MgSO₄·7H₂O 400 mg/l

NH₄H₂PO₄ 300 mg/l

CaCl₂·2H₂O 200 mg/l

25 MnSO₄·H₂O 10 mg/l

H₃BO₃ 5 mg/l

ZnSO₄·7H₂O 1 mg/l

KI 1 mg/l

CuSO₄·5H₂O 0.2 mg/l

	NaMoO ₄ ·2H ₂ O	0.1 mg/l
	CoCl ₂ ·6H ₂ O	0.1 mg/l
	FeSO ₄ ·7H ₂ O	15 mg/l
	Na ₂ EDTA	20 mg/l
5	Schenk and Hildebrant vitamine solution:	
	myo-inositol	1000 mg/l
	Nicotinic acid	5 mg/l
	Thiamine·HCl	5 mg/l
10	Pyridoxine·HCl	0.5 mg/l
	30 g/l saccharose	
	0.5 g/l MES (Sigma)	
	pH= 7.5 (adjusted with KOH)	
	Sterilization: 121 C°, 20 minutes.	
15	The composition of the B5h medium:	
	1 liter of B5 solution	
	30 ml of B5 amino acid stock solution	
	1 ml of B5 hormone stock solution	
20	The composition of the B5 solution:	
	Gamborgs' B5 salts:	
	KNO ₃	2500 mg/l
	MgSO ₄ ·7H ₂ O	250 mg/l
25	(NH ₄) ₂ SO ₄	134 mg/l
	CaCl ₂ ·2H ₂ O	150 mg/l
	NaH ₂ PO ₄ ·H ₂ O	150 mg/l
	MnSO ₄ ·H ₂ O	13.5 mg/l
	H ₃ BO ₃	3 mg/l

	ZnSO ₄ ·7H ₂ O	2 mg/l
	KI	0.75 mg/l
	CuSO ₄ ·5H ₂ O	0.025 mg/l
	NaMoO ₄ ·2H ₂ O	0.25 mg/l
5	CoCl ₂ ·6H ₂ O	0.025 mg/l
	FeSO ₄ ·7H ₂ O	27.8 mg/l
	Na ₂ EDTA	37.3 mg/l
	Gamborgs' B5 vitamine solution:	
10	myo-inositol	100 mg/l
	Nicotinic acid	1 mg/l
	Thiamine·HCl	10 mg/l
	Pyridoxine·HCl	1 mg/l
15	0.5 g/l proline	
	40 g/l saccharose	
	pH= 5.7 (adjusted with KOH)	
	8 g/l Phytoagar (GIBCO BRL)	
20	Sterilization: 121 C°, 20 minutes.	

Before plating sterile amino acid and hormone stock solutions are added.

B5h amino acid stock solution:

25	6.65 g/250 ml L-glutamine
	0.83 g/250 ml serine
	0.004 g/ 250 ml adenine
	0.083 g/250 ml glutathione

Filter sterilization. 30 ml of amino acid stock solution is added to 1 l of B5h medium before plating

5 B5h hormone stock solution:

1 mg/ml 2, 4-D-(2,4-dichlore-phenoxy-acetic acid, GIBCO BRL)

0.1 mg/ml kinetin (GIBCO BRL)

Filter sterilization. 1 ml of hormone stock solution is added to 1 l of B5h medium before plating

The composition of the B5hKmCb medium:

B5h medium containing 25 mg/l of kanamycine and 500 mg/l of carbenicilline.

15

The composition of the MMSCb medium:

4.3 g/l Murashige and Skoog salt (GIBCO BRL)

1 ml/l 1000x Nitsch and Nitsch vitamine stock solution:

	myo-inositol	100 mg/ml
20	glycine	2 mg/ml
	Nicotinic acid	5 mg/ml
	Pyridoxine-HCl	0.5 mg/l
	Thiamine-HCl	0.5 mg/l
	Folic acid	5 mg/l
25	Biotine	0.05 mg/l

30 g/l saccharose

pH= 5.7 (adjusted with KOH)

8 g/l Phytoagar (GIBCO BRL)

Sterilization: 121 C°, 20 minutes.

500 mg/l of carbenicilline is added after sterilization and before plating.

Example 13.The introduction of the NORK gene of *M. truncatula* into other (leguminous and non-leguminous) plants

5

The *Agrobacterium tumefaciens* strains described in Example 12. (BRC1677, BRC1680) carrying the NORK gene of *M. truncatula* were used for the introduction of the NORK gene into the legume plant vetch (*Vicia villosa*) and the non-legume tobacco (*Nicotiana tabacum* cv. Small Havana SR1) plant. The transformations were performed as described in Example 12C with the modification that in the case of *Vicia* 2 mg/l of 2,4-D and 0.2 mg/l of kinetin was used, while the pieces of the tobacco leaves were placed onto MS medium (68) containing 1 mg/l of benzyl-aminopurine (BAP) and 0.1 mg/l of naphthyl-acetic acid (NAA), as well as antibiotics (100 mg/l of kanamycin and 500 mg/l of carbenicillin).

15 After the transformations six *N. tabacum* and eight *V. villosa* plants originated from independent transformation events were regenerated (Table 11.). Total DNA from the transformed plants was isolated and PCR amplification using the Pr-RKU4---Pr-RKD5 primerpair (Figure 9.) was carried out as described in Examples 2. and 9., respectively. The sequence of amplified products (exon1-
20 exon2 of the NORK gene) were determined as described in Example 8. The sequencing results revealed that all sequences determined from the transformed plants were identical to the DNA sequence of the *M. truncatula* NORK gene, which means that the transgenic plants contains the NORK gene of *M. truncatula*. Proteins from the roots of transformed tobacco plants were isolated as described
25 in Example 14. and the presence of the NORK protein in the transformants was demonstrated by dot blot hybridization. These results show that the NORK gene is expressed in *N. tabacum*. The results of the transformation experiments and the characterization of the transformants are shown in Figure 32.

The composition of the MS medium:

Murashige and Skoog salts:

	KNO ₃	1900 mg/l
	MgSO ₄ ·7H ₂ O	70 mg/l
5	NH ₄ NO ₃	650 mg/l
	CaCl ₂ ·2H ₂ O	40 mg/l
	KH ₂ PO ₄ ·H ₂ O	70 mg/l
	MnSO ₄ ·H ₂ O	22.3 mg/l
	H ₃ BO ₃	8.2 mg/l
10	ZnSO ₄ ·7H ₂ O	8.6 mg/l
	KI	0.83 mg/l
	CuSO ₄ ·5H ₂ O	0.025 mg/l
	NaMoO ₄ ·2H ₂ O	0.25 mg/l
	CoCl ₂ ·6H ₂ O	0.025 mg/l
15	FeSO ₄ ·7H ₂ O	27.8 mg/l
	Na ₂ EDTA	37.3 mg/l

Murashige and Skoog vitamine solution:

	myo-inositol	100 mg/l
20	Nicotinic acid	0.5 mg/l
	Thiamine·HCl	0.1 mg/l
	Pyridoxine·HCl	0.1 mg/l
	Glycine	2 g/l
	Saccharose	20 g/l
25	pH= 5.7 (adjusted with KOH)	
	Phytoagar (GIBCO BRL) 8 g/l	

Sterilization: 121 C°, 20 minutes.

Example 14

Immunological detection of the NORK protein

5

In order to prove the presence of NORK protein in the cells of alfalfa, and in the transformed tobacco, a non-legume plant, two exons (2nd and 3rd exons) of the N-terminal (5') part of the NORK gene were expressed in *Escherichia coli*. Antibodies were produced against the purified proteins and the NORK protein was detected by immunological method (dot blot analysis).

Example 14A.

Cloning of the DNA fragment coding the N-terminal part of *M. truncatula*

15 A17 NORK protein into the expression vector pTrcHisA

The pTrcHis (A, B and C) (Invitrogen BV Groningen, The Netherlands) expression vectors are made possible to express recombinant protein containing six histidine (His-tag) and a recognition site of enterokinase enzyme in fusion. The His-tag allows us to purify the recombinant protein in a single-step purification procedure (recommended by the suppliers). For this reasons this system was used to express the extracellular part of the NORK protein.

A DNA fragment of the *M. truncatula* NORK cDNA sequence described in Example 9 and showed in Figure 10 was produced by PCR amplification using the RKEx2U and Pst-A3 primers (Figure 9.) as it is described in Example 9A. The 957 bp long amplified cDNA fragment coded 319 amino acids (amino acids from 33 to 352) of the supposed extracellular part of the NORK protein. pTrcHisA vector was used for cloning to fuse the sequence of the NORK protein portion to be expressed and the His-tag in frame. The pTrcHisA vector and the amplified cDNA fragment were digested by *Bam*HI and *Pst*II restriction enzymes (Fermentas AB) according the supplier's instructions. The fragment and the vector prepared in this way were ligated (construct pTrcHisA::dNORK) as it is described in Example 4. The ligation is transformed into *E. coli* TOP10 competent cells detailed also in

Example 4. After transformation and plating, plasmid DNA was isolated from the ampicillin resistant colonies using the QIAGEN Plasmid Isolation Kit (see Example 4.). The frame and the fidelity of the sequence of the insert were checked by sequence analysis (see Example 8.). The amplified and correctly inserted clone
5 was termed TOP10(pTrcHisA::dNORK).

Example 14B

Expression of the extracellular part of the NORK protein in *E. coli*

10

The recombinant protein (rNORK) produced by the Xpress System Protein Expression system (construct pTrcHisA::dNORK) was purified by the Xpress System Protein Purification (Invitrogen BV Groningen, The Netherlands) according to the manufacturer's instructions. Overnight culture of the clone
15 TOP10(pTrcHisA::dNORK) (see Example 14A.) grown in 2 ml of LB medium containing ampicillin in 50 µg/ml concentrations was diluted 250 times in 50 ml of the same medium. The culture was grown at 37°C with vigorous shaking to an OD₆₀₀=0.6. Then the expression of protein was induced by adding 0.5 ml IPTG solution (100 mM) and the culture was shaken for 5 more hours. Bacterial
20 suspension was collected by centrifugation (centrifuge: Sorvall RC5, GSA rotor with the proper tubes, 4°C, 10 min, 8k rpm) than the bacterial pellet was resuspended in 10 ml of buffer (recommended by the kit: 20 mM NaH₂PO₄, 500 mM NaCl; pH 7.8). Disruption of bacteria was made by sonication 3 times for 10 sec using the smallest head of MSE Ultrasonic Disintegrator. The supernatant of
25 the disrupted and collected bacterial suspension was bound to affinity column through the six histidine amino acid residue present in the fusion protein. Bacterial proteins were removed by three washing steps using solutions with decreasing pH (20 mM NaH₂PO₄, 500 mM NaCl; pH 7.8; 6.0 and 5.5). The recombinant protein was eluted from the column by 5 ml elution buffer of the kit (20 mM NaH₂PO₄, 500
30 mM NaCl; pH 4.0) having 1 mg/ml protein concentration. The eluted fusion protein was used directly to produce antibodies in rabbits.

Example 14CProducing polyclonal antibodies against the extracellular part of the NORK protein

5

Three, 4 months-old rabbits were injected by native protein in 1 ml final volume dissolved in PBS (69) and purified as described in Example 14B, respectively. Approximately 100 µg of total purified protein (antigen) was injected subcutan into four different sites of the neck. For the primary injection 50% complete Freund's adjuvant (CFA) (cat. no: F-5881 Sigma-Aldrich Kft. Budapest) and 50% recombinant protein dissolved in PBS were used. Before the primary injection the rabbits were bled and the gained sera were used later as a negative (isotype) control.

After the fourth and seventh weeks of the primary injection boost injection was performed to achieve a higher state of immunity (hyperimmunization). For the second injection 10% CFA + 40% incomplete Freund's adjuvant (IFA) (cat. no: F-5506 Sigma-Aldrich Kft. Budapest) and 50% recombinant protein dissolved in PBS were used. Ten days later the antibody production was checked. Serum was prepared from 1-5 ml blood obtained from ear's vein. The obtained blood was kept at room temperature for 4 hours and after coagulation it was centrifuged (2700g, 4°C, 10 min). A part of the supernatant (serum) was kept at 4°C until utilization and the rest was stored at -20°C in 100 µl aliquots. The antibody content of the antisera was checked by ELISA (Enzyme-Linked Immunosorbent Assay) and dot blot analysis. The rabbits were kept alive and were injected again on the 7th week after the primary injection. For the third injection 50% IFA and 50% recombinant protein dissolved in PBS were used. On the 10th day after the third injection, the rabbits were bled and the samples were checked again by ELISA and dot blot analysis. Then 80 ml blood was taken from the rabbits.

30

Example 14DELISA

5 500 µg recombinant protein dissolved in PBS per well was plated into 96-well microtiter plate and incubated at 37°C for 2 hours. Then the plate was washed three times by washing buffer (0.05% Tween-20, 0.5 M NaCl in PBS) and was blocked by 0.5% gelatine (1 hr, 37°C). After three washes the plate was incubated at 37°C for 1 hour in the presence of 100 and 1000-fold diluted rabbit
10 anti sera in 50µl washing buffer. After three repeated washes the plate was incubated (1 hr, 37°C) with the second antibodies diluted in 50µl washing buffer for 10000-fold. Horseradish peroxidase conjugated anti-rabbit IgG antibody (cat. No: A-0545 Sigma-Aldrich Kft. Budapest) was used as a secondary antibody. After four washes the plate was developed by 0.5 mg/ml OPD substrate (cat. no: P-
15 5412 Sigma-Aldrich Kft. Budapest). OPD was dissolved in PBS containing 0.1 M citrate than 5 µl 30% H₂O₂ was added to it. After blocking the reaction by 4 M sulphuric acid (50 µl/well), the plates were read with a microplate ELISA Reader (Labsystem Multiscan Biochromatic) at 492 nm. The most specific antibody selected based on the results of ELISA test was used in 1000-fold dilution to
20 investigate the extracts purified from roots of different plants.

Example 14E.

Protein purification from the roots of mutant (Nod⁻) and wild type (Nod⁺)
25 *Medicago sativa* and transformed and control *Nicotiana tabacum* plants

Purification of proteins isolated from the roots of MN-1008 (Nod⁻) and NAB814 (Nod⁺) *Medicago sativa* as well as control (non-transformed) and transformed *Nicotiana tabacum* plants (see Example 13) were performed by the
30 slightly modified methods of Huang and Berry (70) as well as Cheong and Hahn (59). 1 g fresh roots were homogenized in 1 ml homogenisation buffer (25 mM Tris-HCl pH 7.0, 30 mM MgCl₂, 2 mM dithiotreitol (DTT), 0.2 mM

phenylmethylsulfonyl fluoride (PMSF)) in the presence of quartz sand on ice. The homogenate was filtered through four layers of cheesecloth (commercially available) and centrifuged at 6000g (4°C, 15 min). Supernatant and the crude membrane fraction obtained by centrifugation of the supernatant (20000g, 4°C, 30 min) were used for dot blot analysis.

Example 14F

Dot blot analysis of proteins purified from the roots of *Medicago sativa* and *Nicotiana tabacum* using specific antibodies against the NORK protein

Ten µg of protein samples purified from alfalfa and tobacco plants (the filtered and centrifuged supernatant of plant homogenized samples described in Example 14E) were dot-blotted onto nitrocellulose filter (Hybond-ECL cat. no: RPN303D, Amersham Pharmacia Biotech Inc, Little Chalfont, Buckinghamshire, England) (Figure. 32.). Dot blot analysis was performed with the 1000x diluted rabbit antisera as detailed below. Filters were blocked by blocking solution (2.5% milk powder, 1% polyvinyl-pyrrolidone (PVP), 0.1% Tween-20 in PBS) at room temperature for 1 hour. The filters were incubated in the same solution containing the generated antibodies at room temperature for another 1 hour. The filters were washed three times each for ten minutes in 50 ml washing solution (0.1% Tween-20 in PBS). The filters were incubated further with the horseradish peroxidase-conjugated anti-rabbit IgG (5000x diluted in washing solution) at room temperature for 1 hour. After three washing steps the filters were developed by chemiluminescent reaction. After incubation of the filters in developing solution (100 mM Tris pH 8.5, 450 µM coumaric acid, 2.3 mM luminol and 0.03% H₂O₂) for 1 min, they were visualized by an X-ray film exposure (Sterling Diagnostic Imaging Inc., Newark, DE 19714 USA) (Figure 32.). The positive signal detected in the case of the Nod⁺ and Nod⁻ alfalfa as well as of the transformed tobacco plant containing the NORK gene demonstrated the reaction of the NORK protein with the produced specific antibody.

Literature references:

1. Kiss, G.B. 1986. A szimbiotikus nitrogénkötés fokozásának lehetőségei. *Növénytermelés* 35:351-361.
2. Baldani, J. I., Caruso, L., Baldani, V. L. D., Goi, S. R., Dobereiner, J. 1997. Recent advances in BNF with non-legume plants. *Soil Biology & Biochemistry* 29: 911-922.
3. Pueppke, S. G. and Broughton, W. J. 1999. *Rhizobium* sp. Strain NGR234 and *R. fredii* USD257 share exceptionally broad, nested host range. *Mol. Plant Microb. Interact.* 12:293-318.
- 10 4. Norris, D.O. 1956. Legumes and the *Rhizobium* symbiosis. *Emp. J. Exp. Agric.* 24:247-270.
5. Schultze, M., Kondorosi, E., Ratet, P., Buiré, M., Kondorosi, A. 1994. Cell and molecular biology of *Rhizobium*-plant interaction. *Int. Rev. Cytology* 156:1-75.
- 15 6. Horváth, B., Kondorosi, E., John, M. , Schmidt, J., Török, I., Györgypál, Z., Barabás, I., Wieneke, U., Schell, J., Konsorosi, A. 1986. Organization, structure and symbiotic function of *Rhizobium meliloti* nodulation genes determining host specificity for alfalfa. *Cell* 46:335-343.
7. Schultze, M. , Kondorosi, A. 1995. What makes nodulation
20 signals host-specific? *Trends in Microbiology* 3:370-372.
8. Long, S. R. 1989. *Rhizobium*-legume nodulation: life together in the underground. *Cell* 56: 203-214.
9. Györgypál, Z., Iyer, N., Kondorosi, A. 1988. Three regulatory *nodD* alleles of divergent flavonoid-specificity are involved in host-dependent
25 nodulation by *Rhizobium meliloti*. *Mol. Gen. Genet.* 212:85-92.
10. Bergman, K., Gulash-Hofee, M. , Hoverstadt, R. E., Larosiliere, R. C., Ronco, P. G., Su, L. 1988. Physiology os behavior mutants of *Rhizobium meliloti*: evidence for a dual chemotaxis pathway. *J. Bacteriol.* 170: 3249-3254.
- 30 11. Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Prome, J. C. and Denarie, J. 1990. Symbiotic host-specificity of *Rhizobium*-

meliloti is determined by a sulfated and acylated glucosamine oligosaccharide signal. *Nature* 344: 781-784.

12. Horváth, B., Bachem, C.W.B., Schell, J., Kondorosi, A. 1987. Host specific regulation of nodulation genes in *Rhizobium meliloti* is mediated by a plant signal, interacting with the *nodD* gene product. *EMBO J.* 6:841-848.

13. Kondorosi, E., Gyuris, J., Schmidt, J., John, M., Duda, E., Hoffmann, B., Schell, J., Kondorosi, A. 1989. Positive and negative control of *nod* gene expression in *Rhizobium meliloti* is required for optimal nodulation. *EMBO J.* 8:1331-1340.

14. Caetano Annoles, G. and Gresshoff, P.M. 1991. Plant genetic control of nodulation. *Ann. Rev. Microbiol.* 45:345-382.

15. Bhuvanesvari, T.V., Solheim, B. 1985. Root hair deformations in the white clover/*Rhizobium trifolii* symbiosis. *Physiol. Plant.* 63:25-34.

16. Peters, N.K., Frost, J.W., Long, S. 1986. A plant flavon, luteolin, induces expression of *Rhizobium meliloti* nodulation genes. *Science* 233:977-980.

17. Rolfe, B. G. and Gresshoff, P.M. 1988. Genetic analysis of legume nodulation. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 39:297-319.

18. Niebel, A., Gressent, F., Bono, J. J., Ranjeva, R., Cullimore, J. 1999. Recent advances in the study of Nod factor perception and signal transduction. *Biochimie* 81: 669-674.

19. Truchet, G., Barker, D. G., Camut, S., de Billy, F., Vasse, J., Huguet, T. 1989. Alfalfa nodulation in the absence of *Rhizobium*. *Mol. Gen. Genet.* 219:65-68.

20. Freiberg, C., Fellay, R., Bairoch, A., Broughton, W. J., Rosenthal, A., Perret, X., 1997. Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* 387: 394-401.

21. Covitz, P. A., L. S. Smith, Long, S. R. 1998. Expressed sequence tags from a root-hair-enriched *Medicago truncatula* cDNA library. *Plant Physiology* 117: 1325-1332.

22. Gyorgyey, J., Vaubert, D., Jimenez-Zurdo, J. I., Charon, C., Troussard, L., Kondorosi, A., Kondorosi, E. 2000. Analysis of *Medicago truncatula* nodule expressed sequence tags. *Molecular Plant-Microbe Interactions* 13: 62-71.
- 5 23. Downie, J. A. and Walker, S. A. 1999. Plant responses to nodulation factors. *Current Opinion in Plant Biology* 2: 483-489.
24. Csanádi, G., Szécsi, J., Kaló, P., Kiss, P., Endre, G., Kondorosi, A., Kondorosi, E., Kiss, G.B. 1994. ENOD12, an early nodulin gene, is not required for nodule formation and efficient nitrogen fixation.
10 *Plant Cell* 6: 201-213.
25. Schauser, L., A. Roussis, Stiller, J., Stougaard, J. 1999. A plant regulator controlling development of symbiotic root nodules. *Nature* 402:191-195.
26. Tanksley, S.D., Ganai, M. W., and Martin, G.B. 1995.
15 *Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends in Genetics*: 11: 63-68.
27. Peterson, M. A. and Barnes, D.K. 1981. Inheritance of ineffective nodulation and non-nodulation traits in alfalfa. *Crop Science* 21:611-616.
- 20 28. Barnes, D. K., Vance, C.P., Heichel, G.H., Peterson, M. A., Ellis, W.R. 1988. Registration of a non-nodulation and three ineffective nodulation alfalfa germplasms. *Crop Sci.* 28:721-722
29. Catoria, R., Galera, C., de Billy, F., Penmetsa, V., Journet, E-P., Maillet, F., Rosenberg, C., Cook, D., Gough, C., Denarie, J. 2000. Four
25 genes of *Medicago truncatula* controlling components of a Nod factor transduction pathway. *The Plant Cell* 12:1647-1665.
30. Walker, S.A., Viprey, V., and J. Downie, A.J. 2000. Dissection of nodulation signaling using pea mutants defective for calcium spiking induced by Nod factors and chitinoligomers. *PNAS* 97: 13413-13418.
- 30 31. Parniske, M. , Coomber, S., Kistner, C., Mulder, L., Pitzschke, A., Stougaard, J., Szczylowski, K., Webb, J., Stracke, S. *Plant Genetics of*

symbiosis. In Abstract book of the EU conference of Molecular Genetics of Model Legumes, JIC, Norwich, June 24-28, 2000.

32. Dudley, M. E., and Long, S.R. 1989. A non-nodulating alfalfa mutant display neither root hair curling nor early cell division in response to
5 *Rhizobium meliloti*. *Plant Cell* 1:65-72.

33. Ehrhardt, D. W., Wais, R., Long, S. R. 1996. Calcium spiking in plant root hairs responding to *Rhizobium* nodulation signals. *Cell* 85: 673-681.

34. Ambudkar, S. V., Sauna, Z. E., Muller, M. M., Kerr, K. M.,
10 Smith, M. M. 2000. Biochemical mechanism of action of multidrug resistance-linked ABC transporters. *Chemistry and Physics of Lipids* 107:1-15.

35. Schwiebert, E. M., Benos, D. J., Fuller, C. M. 1998. Cystic fibrosis: A multiple exocrinopathy caused by dysfunctions in a multifunctional
15 transport protein. *American Journal of Medicine* 104: 576-590.

36. Song, W. Y., Wang, G. L., Chen, L. L., Kim, H. S., Pi, L. Y., Holsten, T., Gardner, J., Wang, B., Zhai, W. X., Zhu, L. H., Fauquet, C., Ronald, P. 1995. A Receptor Kinase-Like Protein Encoded by the Rice Disease Resistance Gene, *Xa21*. *Science* 270: 1804-1806.

20 37. Li, J. M. and Chory, J. 1997. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* 90 929-938.

38. Trotochaud, A. E., Hao, T., Wu, G., Yang, Z., Clark, S. E. 1999. The *CLAVATA1* receptor-like kinase requires *CLAVATA3* for its assembly
25 into a signaling complex that includes KAPP and a rho-related protein. *Plant Cell* 11: 393-405.

39. Kiss, G.B., Kereszt, A., Kiss, P., Endre, G. 1998. Colormapping: a non-mathematical procedure for genetic mapping. *Acta Biologica Hungarica* 49:47-64.

30 40. Kaló, P., Endre, G., Zimányi, L., Csanádi, G., Kiss, G.B. 2000. Construction of an improved linkage map of diploid alfalfa (*Medicago sativa*). *Theor. Appl. Genet.* 100:641-657.

41. Gibson, A.H. 1980. Methods for legumes in glasshouses and controlled environment cabinets. In: Methods for evaluating biological Nitrogen Fixation. Bergersen, F.J. ed., John Wiley and Sons Ltd. p.139-184.
42. Szende, K, und Ördögh, F 1960. Die Lysogenie von *Rhizobium meliloti*. *Naturwissenschaften* 47:404-405.
43. De Lajudie, P., Willems, A., Pot, B., Dewettinck, D., Maestrojuan, G., Neyra, M. , Collins, M. D., Dreyfus, B., Kersters, K., and Gillis, M. 1994. Polyphasic taxonomy of rhizobia: Emendation of the genus *Sinorhizobium* and description of *Sinorhizobium comb. nov.*, *Sinorhizobium saheli* sp. nov. and *Sinorhizobium teranga* sp. nov. *Int. Journal of Systemic Bacteriol.* 44:715-733.
44. Orosz L, Sváb Z, Kondorosi Á, Sík T 1973. Genetic studies on rhizobiophage 16-3. I. Genes and functions on the chromosome. *Mol Gen Genet* 125:341-350.
45. Michelmore, R.W., Paran, I., and Kesseli, R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. USA*, 88:9828-9832.
46. Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. and Tingey, S.V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18:6531-6535.
47. Kiss, G.B., Csanádi, G., Kálmán, K., Kaló, P., Ökrész, L. 1993. Construction of a basic genetic map of *Medicago* using RFLP, RAPD, Isozyme and morphological markers. *Mol Gen Genet* 238:129-137.
48. Church, G. M. and Gilbert, W. 1984. Genomic sequencing. *PNAS* 81: 1991-1995.
49. Maniatis, T., Fritsch, E.F., Sambrook, J. 1982. *Molecular Cloning; A Laboratory Manual*; Cold Spring Harbor Laboratory. 447. oldal.
50. Vieira J., Messing J. 1991. New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene* 100:189-194.

51. Turner SR, Ireland R, Morgan C, Rawsthorne S 1992. Identification and localization of multiple forms of serine hydroxymethyltransferase in pea (*Pisum sativum*) and characterization of a cDNA encoding a mitochondrial isoform. *Journal of Biol Chem* 267:13528-13534.
52. Woo, S. S., Jiang, J. M. , Gill, B. S., Paterson, A. H. and Wing, R. A. 1994. Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Research* 22: 4922-4931.
53. Nam YW, Penmetsa RV, Endre G, Uribe P, Kim DJ, Cook DR 1999. Construction of a bacterial artificial chromosome library of *Medicago truncatula* and identification of clones containing ethylene response genes. *Theor. Appl. Genet.* 98:638-646.
54. Koncz, C., and Rédei, G. 2000. A molekuláris genetika eszközei: *Agrobacterium* T-DNS és *Arabidopsis*. Molekuláris növénybiológia C. Könyvben, szerkesztők: Balázs E., Dudits D. Akadémiai Kiadó. 323-452. oldal.
55. Quandt, H.-J., Pühler A., and Broer I. 1993. Transgenic root nodules of *Vicia hirsuta*: A fast and efficient system for the study of gene expression in indeterminate-type nodules. *Mol. Plant-Microbe Interact.* 6:699-706.
56. Stougaard, J. 1995. *Agrobacterium rhizogenes* as a vector for transforming higher plants. Application in *Lotus corniculatus* transformation. *Methods Mol. Biol.* 49:49-61.
57. Bingham, E. T., Hurley, L. V., Kaatz, D. M., and Saunders, J. W. 1975. Breeding alfalfa which regenerates from callus tissue in culture. *Crop Sci.* 15:719-721.
58. Brosius, J. 1989. Superpolylinkers in cloning and expression vectors. *DNA.* 8:759-777.
59. An, G., Watson, B. D., Stachel, S., Gordon, M. P., Nester, E.W. 1985. New cloning vehicles for transformation of higher plants. *EMBO J.* 4:277-284.

60. Szabados, L., Charrier, B., Kondorosi, A., de Bruijn, F., Ratet, P. 1995. New plant promoter and enhancer testing vectors. *Mol. Breeding* 1:419-423.
61. Hellens, R.P., Edwards, E. A., Leyland, N. R., Bean, S. Mullineaux, P.M. 2000. pGreen: a versatile and flexible binary Ti vector for Agrobacterium-mediated plant transformation. *Plant Mol. Biol.* 42: 819-832.
62. Vancanneyt, G., Schmidt, R., O'Connor-Sanchez, A., Willmitzer, L., Rocha-Sosa, M. 1990. Construction of an intron containing marker gene: Splicing of the intron in transgenic plants and its use in monitoring early events in Agrobacterium-mediated transformation. *Mol. Gen. Genet.* 220:245-250.
63. Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M., Peterson, K. M. 1994. pBBR1MCS: a broad-host-range cloning vector. *Biotechniques*.16:800-802.
64. Pridmore, R. D. 1987. New and versatile cloning vectors with kanamycin-resistance marker. *Gene* 56:309-312.
65. Sharma, R. C., and Schimke, R. T. 1996. Preparation of electrocompetent E. coli using salt-free growth medium. *Biotechniques*. 20:42-4.
66. Shahin, E. A. 1985. Totipotency of tomato protoplasts. *Theor. Appl. Genet.* 69:235-240.
67. Jefferson, R. A. 1989. The GUS reporter gene system. *Nature* 342:837-838.
68. Murashige, T., and Skoog, F. 1962. A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* 15:473-497.
69. Coico, R., Coligan, J.E., Kruisbeek, A.M., Margulies, D.H., Shevach, E.M., and Strober, W. *Current Protocols in Immunology* 1994. John Wiley & Sons Inc. USA.
70. Huang, L. S. and Berry, E. A. 1990. Purification and characterization of the proton translocating plasma membrane ATPase of red beet storage tissue. *Biochim Biophys Acta* 1039:241-52.

71. Cheong, J. J., and Hahn, M. G. 1991. A specific, high-affinity binding site for the hepta-beta-glucoside elicitor exists in soybean membranes. *Plant Cell* 3:137-47.

CLAIMS

1. Nucleotides sequence coding for the polypeptide having NOD region specific receptor kinase (NORK) gene activity.
- 5 2. Nucleotide sequence coding for the amino acid sequence of Figure 17.
3. Nucleotide sequence coding for the amino acid sequence of Figure 12.
4. Nucleotide sequence coding for the amino acid sequence of Figure 30.
5. Nucleotide sequence coding for the amino acid sequence of Figure 31.
6. The nucleotide sequence according to any of claims 1 to 5 which is a
- 10 genomic or synthetic DNA molecule:

CCATATTTTAACAATATTCTTTCTTCTACAAGGGTATAACTTTTATACAAGTTCACTATA
TTATAGGATTGATCAAGGTTCAATTTTTCTTTCTTTGAAAAATCTCTAAGGGGTGTGGTT
TCCAAGGCAGAAAATGAAATAGAATGCAGAAGAATTTGTATGGTACTATAAAGGGAAGAT
15 GAAAAGTTAGTTAGCATGGATTCAAGTTTGATAACCCCTTTGGGGTAAAATCTCTTTCAGA
TTATGATGGAGCTACAAGTTATTAGGATATTTAGATTGGTTGTGGCATGTGTTCTTTGTT
TGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAGGGTTTGAGAGCATATCATGTT
GTGCTGATTCCAATTACACAGATCCAAAAACAACCTTAACCTTATACAACAGATCACATCT
GGTTCCTCTGATAAAAGAAGTTGCAGACCAATACCCGAAATTTTGTGTTAGCCACAGAAGCA
20 ATAAAAATGTTTGAATATTTGAAATAGATGAAGGAAAGAGATGTTATACTTTGCCAACAA
TTAAGGATCAAGTATATTTGATAAGGGGTGTATTTCCCTTTGATAGTTTAAATTCCTTCGT
TTTATGTTTATATCGGGGTAAACAGAACTAGGTGAATTAAGATCGTCTAGACTCGAGGACT
TGGAATCGAGGGAGTTTTTTAGAGCCACCAAAGACTATATTGATTTCTGCTTATTGAAGG
AAGATGTCAATCCCTTCATTTCTCAGATTGAATTGAGGCCATTACCTGAAGAATACCTAC
25 ATGGTTTTCGCTACTAGTGTTTTAAAACTGATAAGCAGAAATAATCTTGGTGACACAAATG
ATGATATAAGGTTCCCAGATGACCAAAATGATAGAATCTGGAAACGGAAAGCAACTTCAA
CTCCATCATCTGCCCTTCCCCTGTCTTCCAATGTCAGCAATGTTGACCTCAAAGACAGTG
TCACACCTCCTCTACAAGTCCTACAAACAGCTCTTACTCACCTGAGCGATTGGAGTTTCG
TCCATGATGGCCTCGAGACCGATGATTATGAATACTCTGTGTTTCTCCACTTTCTTGAAC
30 TAAATGGCACTGTCAGAGCAGGACAAAGGGTGTGTTGACATCTATCTAAACAATGAGATTA
AAAAGGAGAAGTTTGATGTTTTGGCTGGAGGGTCCAAGAACAGTTACACTGCCCTTGAACA
TTTCAGCAAATGGATCACTCAATATAACCTTAGTCAAGGCATCTGGATCTGAGTTTGGAC
CCCTTTTGAATGCCTATGAAATCCTGACGGCACGGTCGTGGATTGAAGAGACCAACCAAA
AAGATTTGGAACCTTATTCAGAAGATGAGAGAAGAACTGCTGCTGCACAACCGAGAAAATG
35 AAGCATTGGAGAGTTGGAGTGGAGACCCCTGTATGATTTTCCCCTGGAAAGGAATAACAT
GTGATGATTCAACTGGTTCATCTATTATCACTATGCTGGATCTTTCTTCCAATAATCTCA
AGGGAGCAATTCCTTACTTTGTCACTAAGATGACCAATTTACAAATACTGAACCTGAGCC
ACAACCAGTTCGATTTCGTTATTCCCCTCGTTTCCACCGTCCTCCTTGCTGATATCATTTGG

ATCTGAGCTACAATGATCTTGATGGACGGCTTCCAGAATCCATTATCTCACTGCCACATT
TAAAATCATTATATTTTTGGCTGCAATCCATATATGAAGGACGAAGATACAACAAAGTTGA
ACAGTTCACTAATCAATACAGATTATGGGAGATGCAAAGGAAAAAACCAGTTTGGAC
AAGTATTCGTGATTGGAGCTATTACAAGGGGATCACTTTTGATTACTTTGGCTGTTGGAA
5 TTCTATTTTTTTGCCGTTATAGACACAAGTCAATTACTTTGGAAGGATTGTTGGTGGAAAGA
CCTACCCAATGGCAACAAATATAATCTTCTCTTTGCCAAGCAAAGACGATTTCTTCATAA
AGTCTGTATCAGTTAAACCATTCACTTTGGAGTATATAGAGCAGGCTACAGAACAGTACA
AACTTTGATAGGTGAAGGAGGATTTGGTTCTGTTTACAGAGGCACTCTAGACGATGGTC
AAGAAGTGGCAGTGAAAGTGGGTCATCCACATCAACTCAGGGAACCCNAGAATTTGATA
10 ATGAGCTAAACCTACTTTCAGCTATACAACATGAGAACCTGGTGCCTCTTCTGGGTTACT
GTAATGAGTATGATCAACAAATCTCTGTGTATCCATTCATGTCCAATGGCTCTTTGCTAG
ATAGACTATACGGGAAGCATCAAAGAGAAAAATATTAGACTGGCCAACTAGACTCTCTA
TTGCTCTCGGTGCAGCTCGAGGTTTGGCATATCTTCACACATTTCCAGGACGTTCTGTAA
TACACAGGGACGTAAAATCGAGCAATATACTGCTGGATCAGAGCATGTGTGCTAAGGTTG
15 CAGATTTTGGTTTCTCAAAATACGCTCCTCAGGAAGGAGACAGTTATGTTTCCCTTGAAG
TAAGAGGAACTGCAGGGTATCTGGATCCTGAGTACTACAAAACCCAGCAATTATCTGAAA
AAAGTGATGTTTTTCAGCTTTGGTGTGGTTCTACTTGAAATTGTAAGTGGACGGGAACCTC
TCAACATAAAGAGACCACGGATCGAGTGGAGCTTGGTTGAATGGGCTAAACCATACATAA
GAGCATCAAAGGTGGATGAAATTGTAGATCCTGGCATCAAGGGAGGATATCATGCAGAAG
20 CATTGTGGAGAGTTGTGGAAGTAGCACTGCAATGTCTAGAACCCTACTCAACATATAGGC
CATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTCATTATTGAAAACAATGCAT
CTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAACCGCTACTCAATTGTTATGG
ACAAACGGGCGCTGCCTTCAACTACATCTACAGCAGAATCAACTATCACAACCCAAACCT
TGACACACCCTCAACCGAGATAGTAAATGGGTCGATGGAATTCCTTTTGATTTGTTTTTTA
25 TCATTGCTTTTAGTAATATCCCATTTTAAATGGTAAAGGAGAAAAATACTACTTTTGATTG
TATTTTCATCCACTCTATGTTTCTTGAACTGAATCTCTCTTGCTCAGCCCCAGTTTTTTA
TGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAAACCATATGGTGCATAATTTGAAA
GCCATATTATATCATTTGCTAAGTCCAAAGTAAAAATTTACAAACTAGTTAGATTGCGA
TTTAGTCTATACACACTTCAACAGAGCTATATACACTAT
30

or a deletion product, recombinant form and variant thereof.

7. The recombinant form of any of the nucleotide sequences of claims 1 to 5 which comprise a functional promoter sequence.

8. Oligonucleotide which comprises at least the functional part of the
35 nucleotide sequence of any of claims 1 to 7.

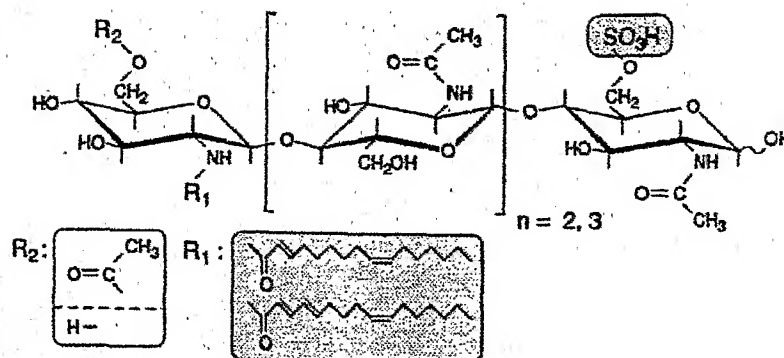
9. Oligonucleotide hybridizing to any of the nucleotide sequences of claims
1 to 8.

10. Antibody capable of detecting the polypeptide having NORK biological activity.
11. Antibody capable of detecting the polypeptide encoded by the DNA sequences according to any of claims 2 to 9.
- 5 12. Transgenic plant which comprises the nucleotide sequence of any of claims 1 to 9.
13. The transgenic plant according to claim 12 which is a crop plant.
14. The transgenic plant according to claim 12 which is tobacco plant.
15. Transformant cell which contains the nucleotide sequence according to
10 any of claims 1 to 9.
16. The amino acid sequence according to Figure 17 and a functional fragment thereof having NORK activity.
17. The amino acid sequence according to Figure 12 and a functional fragment thereof having NORK activity.
- 15 18. The amino acid sequence according to Figure 30 and a functional fragment thereof having NORK activity.
19. The amino acid sequence according to Figure 31 and a functional fragment thereof having NORK activity.
- 20 20. Process for preparing plants of Nod⁺ phenotype from plants of Nod⁻ phenotype characterized by transforming a plant or the reproducing form thereof with a DNA according to any of claims 1 to 9.

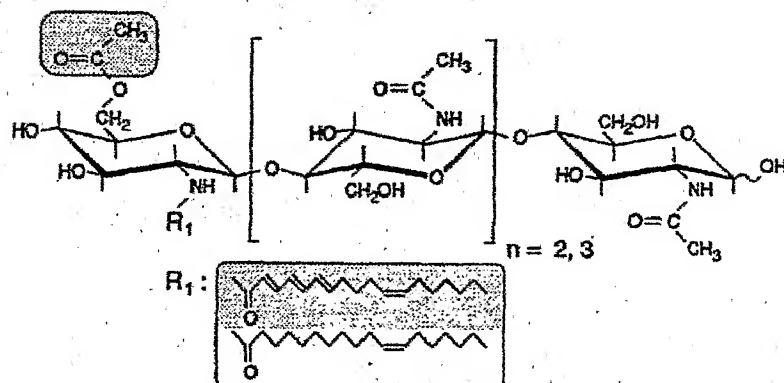
1/60

Figure 1. (1/2)

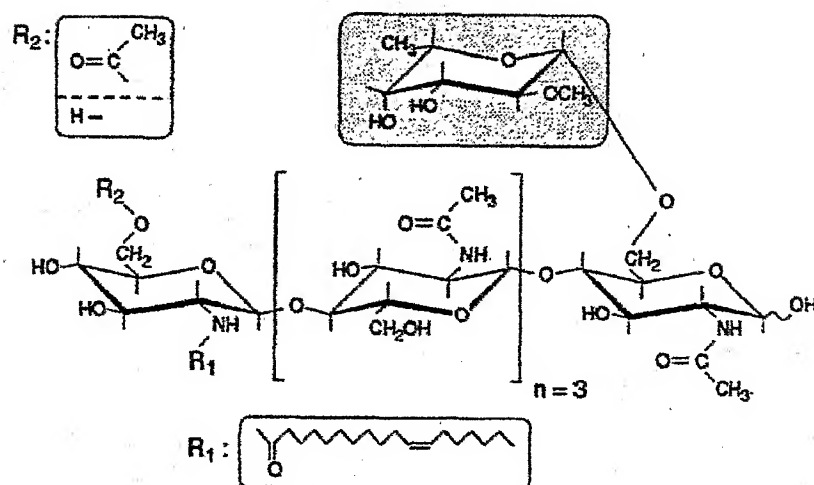
A.



B.

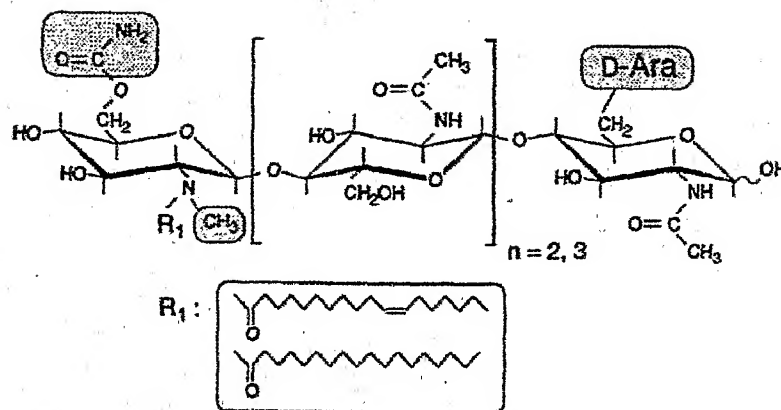
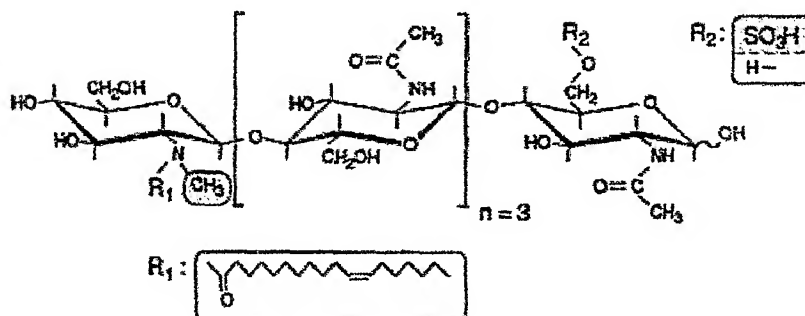
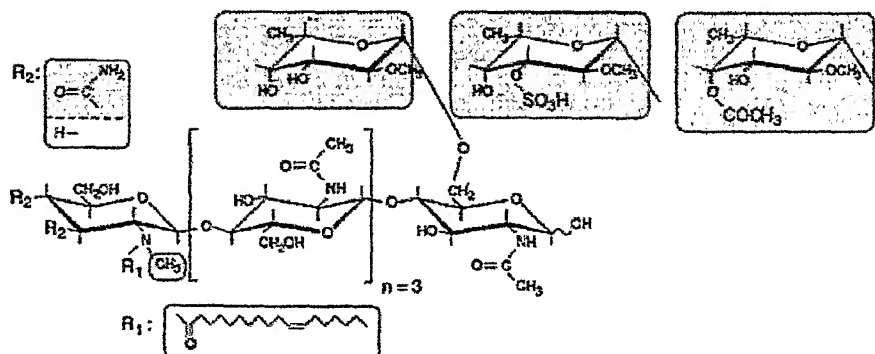


C.



2/60

Figure 1. (cont. 2/2).

D.**E.****F.**

3/60

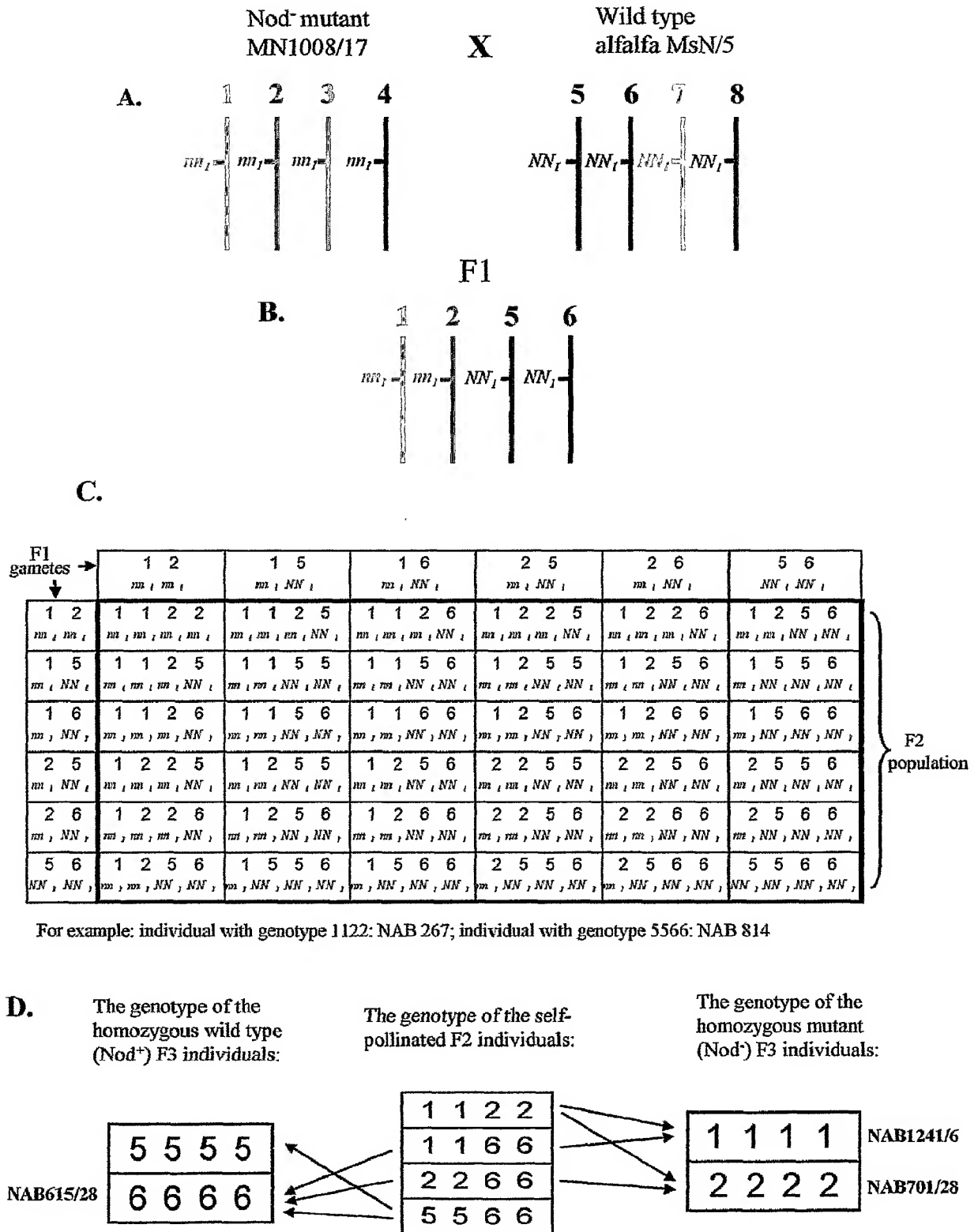
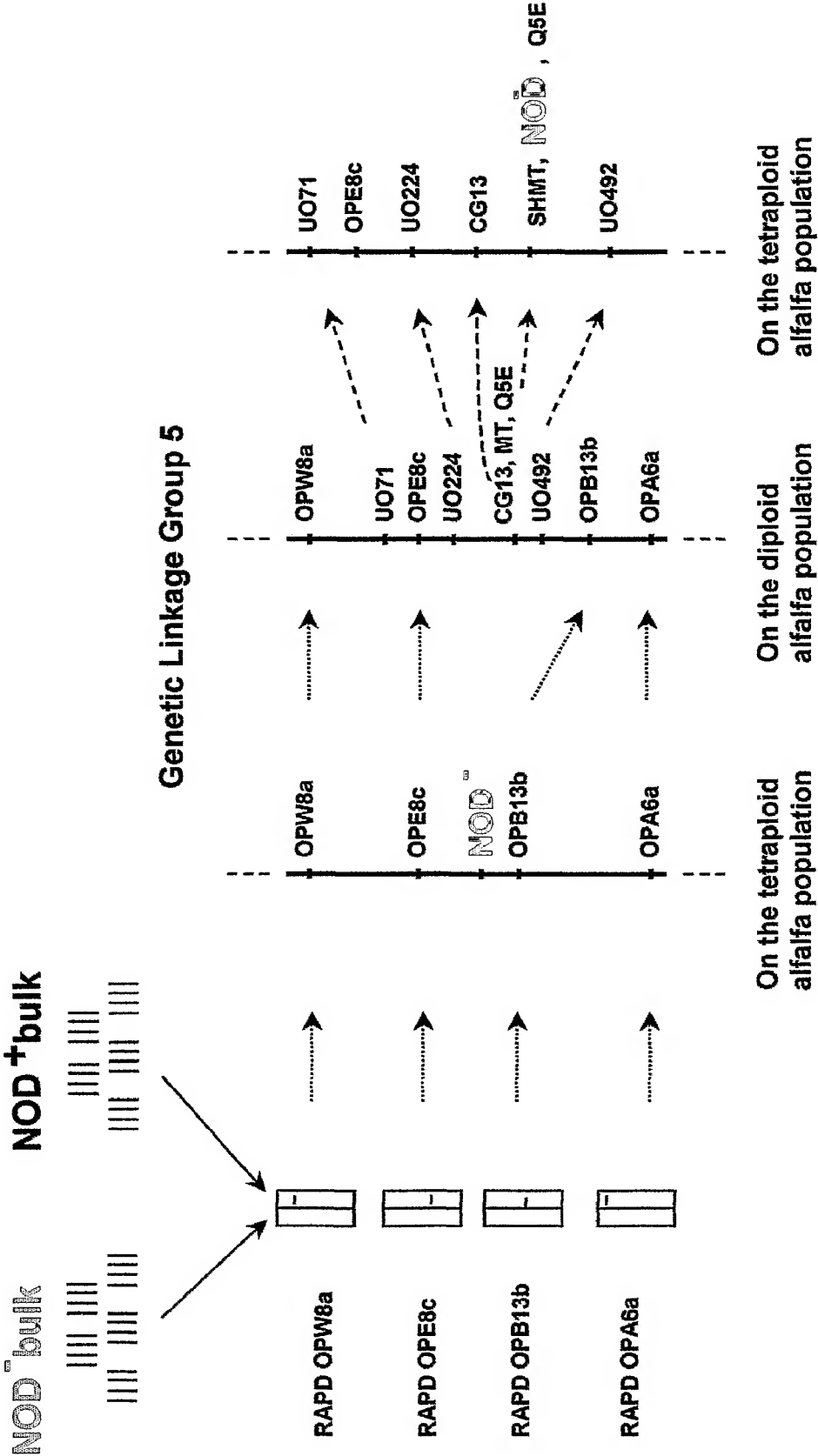
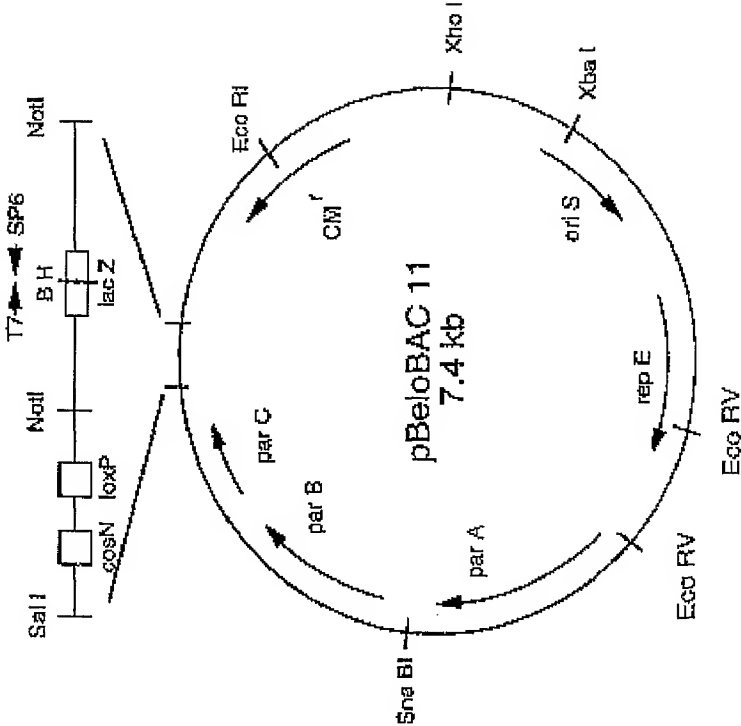
Figure 2.

Figure 3.



||||| ||||| = The 4 homologous chromosome segments representing the Nod region of the tetraploid Nod⁻ and Nod⁺ individuals

Figure 4.



6/60

Figure 5.

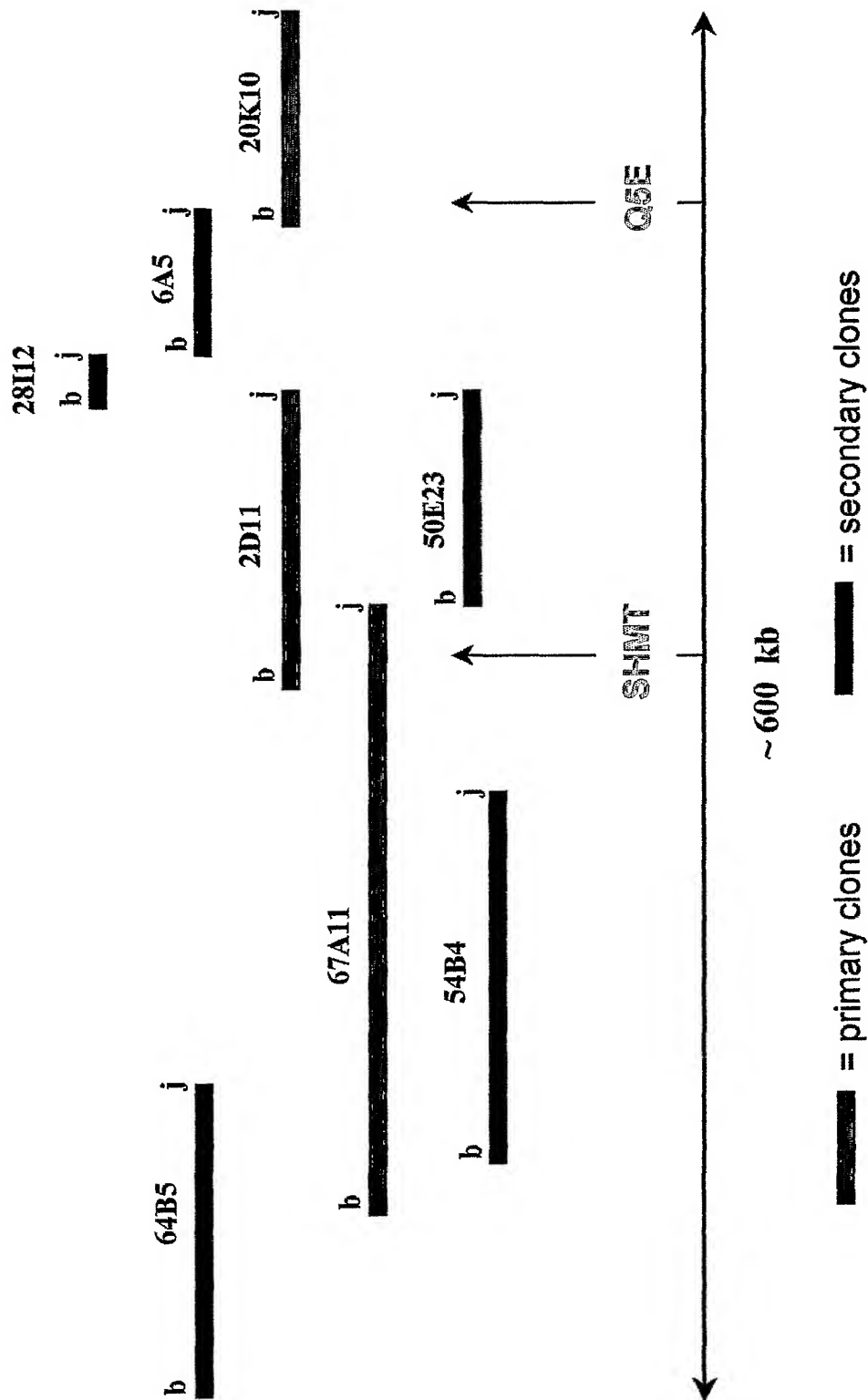


Figure 6.

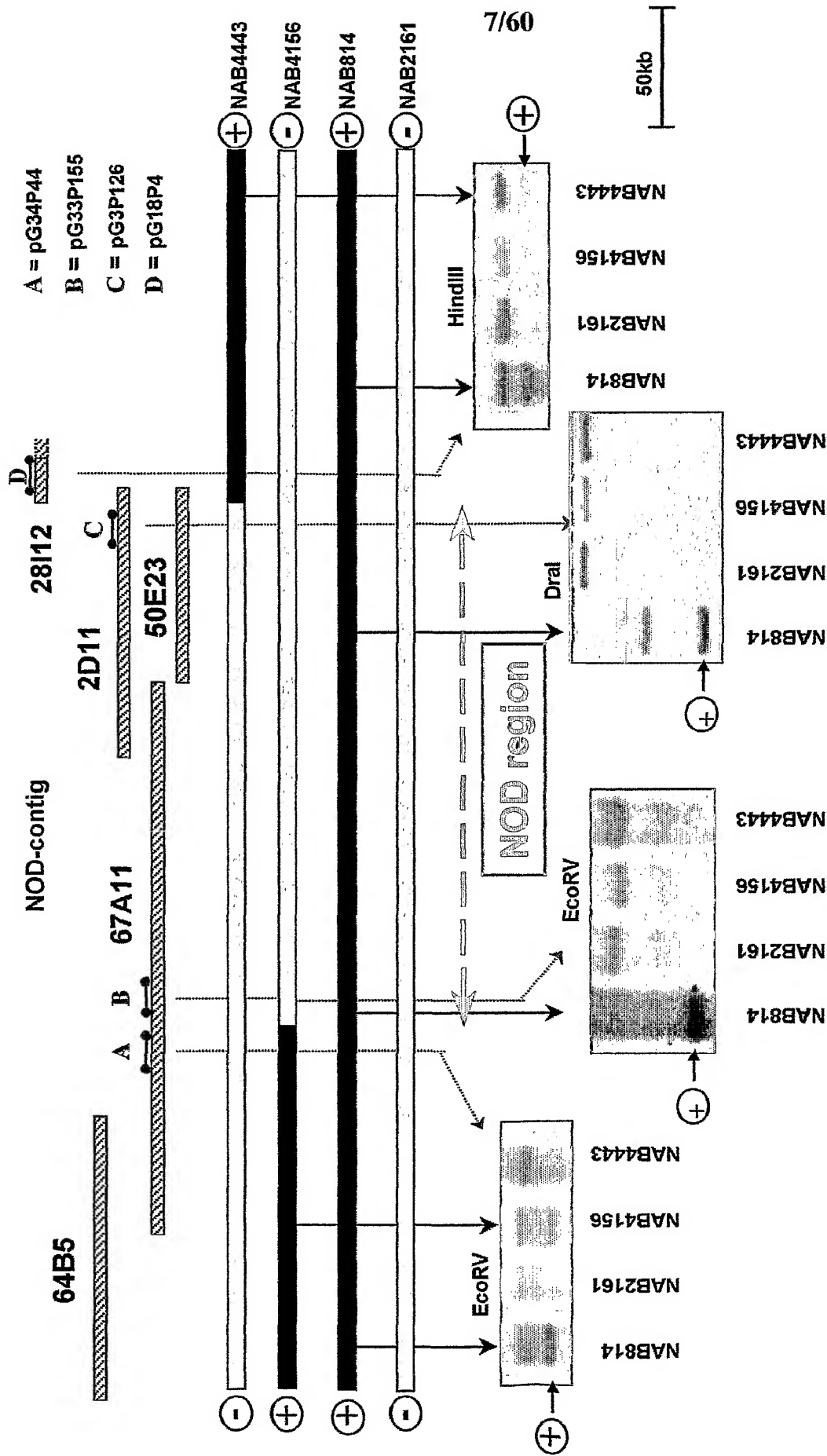
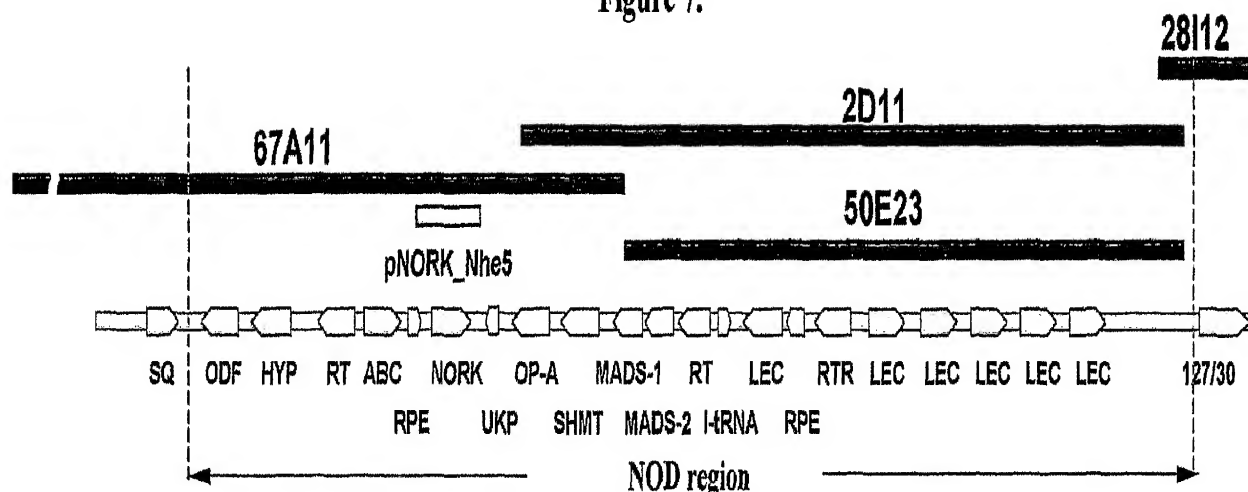


Figure 7.



Abbreviation	Name of the gene	Accession number	BLAST score	E value
SQ:	Coding for squallen epoxidase	dbj/BAA24448.1	148	2e-34
ODF:	Expressed in mouse sperm	gb/AAB54209.1	96.3	2e-18
HYP:	Coding for hypothetical protein	gb/AAD15574.1	185	3e-45
RT:	Coding for reverse transcriptase	emb/CAA73798.1	66	7e-10
ABC:	Coding for ABC transporter	emb/CAB82705.1	131	4e-29
RPE:	Repetitive DNA	gb/AAB50037.1	56.6	6e-07
NORK:	Coding for receptor kinase (NORK)	dbj/BAB09508.1	82.3	1e-13
UKP:	Coding for unknown protein	gb/AAD15572.1	88	4e-17
OP-A:	Coding for oligopeptidaz A	dbj/BAA98181.1	208	9e-62
SHMT:	Coding for serine hidroxymethyl transpherase	gb/AAA33687.1	139	3e-34
MADS-1:	Coding for MADS-box protein	emb/CAA71739.1	51.5	1e-05
MADS-2:	Coding for MADS-box protein	emb/CAA71739.1	51.5	1e-05
RT:	Coding for reverse transcriptase	gb/AAD21515.1	34.8	6.2
I-tRNS:	Coding for isoleucine tRNA	dbj/D82067.1	129	1e-28
LEC:	Coding for lectin precursor	gb/AAB51442.1	127	1e-35
RPE:	Repetitive DNA	gb/AAB50037.1	122	2e-26
RTR:	Retrotransposon	dbj BAA96774.2	201	3e-50
LEC:	Coding for lectin precursor	gb/AAB51442.1	84.3	1e-38
LEC:	Coding for lectin precursor	gb/AAG00508.1	96.7	3e-36
LEC:	Coding for lectin precursor	gb/AAB51442.1	132	5e-30
LEC:	Coding for lectin precursor	gb/AAB51442.1	137	1e-31
LEC:	Coding for lectin precursor	gb/AAB51442.1	138	8e-32
127/30:	LG127/30 like gene	emb/CAB37543.1	82.3	7e-22

9/60

Figure 8. (1/3)

GCTAGCTTAATTTGTGAAGTAGTTTTTCAGTCAGCTCATGCTGAACAGTTATGGTAGTTTT
ACAAGGATGTAAGGGGTTACTTGCTCTTTTTGTGGTATCATTGAGCCACTTTCCACTTTCT
TAATTCCTTTGCTATCAAATTTAAATATATTATATTATTGTAGGAAATATGAAAAAGAAAC
CCTAGATGCTTTAATTGTGCTAAGCATATAAGATACAAAAAAAAAAGTCTAATCTCTT
TTGATAGGAACTGACACAAGGTTGATCCTCAGAACCAACTATGAGACTAATTTTTGCT
TGAATTGGTTAATCAACTGCATCGCACTCATTGAATAAAAAAGTACGTTTAGATTGACGG
TGAATTTAACATAATCACAGTTGACCACTTTGATTTTCACGAAAGTTGCATGTAACCTTC
GTGCTTGAATATAAAGTCACGGTGCCACAATGGTTTTGTCAATATCGCCGTGAATCCAAA
CATGCACCTAATATGTTTTTATATTAATGGAAACCTATGTATATAAATATAAAATACCTTA
AGCATATATATCCTTGCAACAATACCAAAAAAGTTTAAAATTTGCTCCAAAGATACTATT
CACGGTTTTCTATTGTGATGCATCCATGGAATGAGAAGGAAGAAAACAGAAGAAGACGAAG
GACTGGTCTGTCCAATTATTTAAACCATATGAATATTTTTATATACAGATAATGAAGCAA
CCACCACTAAAACAATAAGAATAAATAAATAATGTTCAACATCATCAATAATAAGAATAAT
AATGTTTAAACATCAACAATAAACTATAGGAAAAAACATAATCAACTATGCATTTGTACT
AATTCATCTCTAACTCGTCTTCCATCTCTTTCCTAGCTACCTCCTGCAGTTTCCTTTCC
AGGCCTAAAGTCAAACACCATATTTTAAACAATATTCTTTCTTCTACAGCTTTTATACAAG
TTCATATATTATAGGATTGATCAGGGTTCATTTTTCTTTCTTTGAAAAATCTCTAAGG
GGTGCTGTTTTCCAAGGCAGAGAATGAAATAGAATTCAGAAGAATTTTATGTTACTATAA
AGGAAAGATGAAAAGTTAGTTAGCATGGATTCAAGTTTGATAACCCATATGGGGTAAAAATC
TCTTTCAGATTATGATGGAGTTACAAGTTATTAGGATATTTAGATTGGTTGTGGCATTG
TTCTTTGTTTGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAGGTTAGTAGCTAA
ATACTATAATTCTTTAAGATCATAATAATATCTATTACTTGATTCTTTCCCTTTAAACA
TAGAAAACAACATATTTTAATTAACATGAAAGGCCATGGGATGATCATAATTAATAATGA
GAAGAAAATAAATAAGAACTTGTTTTTTTTTACAAGGATTGTAAAATAGAAGTAGTAGTT
TCATACTTTCAATACTGAGAATCTTGAAACAATTTCACTTTTTCTTTATGTTGCTAGAAT
TTCTTTCAAGGGAAAATCCAATTTTGTACAAAATGAATTTAACTTGTGACATTTTCCTTG
TAGGGTTTGAGAGCATAGCATGTTGTGCTGATTCAAATTACACAGATCCAAAAACCACCC
TAACTTATACAACAGATCACATCTGGTTCTCTGATAAAAGAAGTTGCAGACAAATACCCG
AAATTTTGTTTAGCCACAGAAGCAATAAAAAATGTTGAAAATTTGAAATATATGAAGGAA
AGAGATGTTATAATTTGCCAACAGTTAAGGATCAAGTATATTTGATAAGGGGCATATTTT
CCTTTGATAGTTTAAATTCCTTCGTTTTATGTTTCGATCGGGGTAACAGAAGTAGGCGAAT
TAAGATCGTCTAGGCTCGAGGACTTGGAATTTAGGGGAGTTTTTAGAGCCACCAAAGACT
ACATAGATTTCTGCTTATTGAAGGAGGATGTCAATCCCTTCATTTCTCAGATTGAATTGA
GGCCATTACCTGAAGAATACCTACATGGTTTCGGTACTAGTGTTTTAAACTGATAAGCA
GAAACAATCTTGGTGACACAAATGATGATATAAGGTATGTGATCTTACTTTATTTTAGG
TAGATTCCACCTCTATTTTACAGGGAGTGTCTCTCAGGAAACCTAAAAGGCTTAGGGTTT
GTCACTACTTGGTTCTTATGGAAGCATCATATGTTCACTTAGTAGATATATATTGTAT
AAAAATCAACATCTTTCTCCGTAAGGGATAGAAATTGCAATTCATGTTATGTTGAGCATA
TTTAAACCAAATAATTTGAGGAATATGATGCATGAACCTTTCTGCCAAATGCATGACAT
AACCTATGTTTCACTTTACATATAGAATTATAAGATGTGTTTACATCTTTATATTAACCTT
TGATCTTTATGATGCACTGAGATACACCTGAATTTAAATAACAAATGGAGAAGCAAGAAT
CTAGAAGCTTACCATTCAATCCATTGTAATATGGAAGTCAACAAGAATTAAGTATTGCCAC
TTCAGATTAAATACCCTAAGAGTGTGTTGTGGATGAGGGAATGTTTTGAGGGAATGTAAT
GTTTTGAGGGAATTCCTCAAAACATTACATTACCTCAAAATATTCCCCATCCAAACACA
CTATAAAGGATCCTAATCTAGTGTAGCCTTAGTTTTCTATTGAAGTGATTTGTGTTTAGA
GTTTGGTCTGCATCTGCAACTTTGTTTTGTGCTGTAATGCCATGACTATGAACACTGAC
AGAAAGGTCTCATAATATCGGTATCTATCAATTTTAGGTTCCAGATGACCAAAATGATA
GAATCTGGAACGGAAAGAACTTCAACTCCAACATCTGCCCTTCCACTGTCTTTCAATG
TCAGCAATGTTGACCTCAAAGACAGTGTACACCTCCTCTACAAGTCTTACAAACAGCTC

10/60

Figure 8. (cont. 2/3)

TTACTCACCCCTGAGCGATTGGAGTTCGTCCATGATGGCCTCGAGACCGATGATTATGAAT
ACTCTGTGTTTTCTCCACTTTCTTGAACATAAATGGCACTGTCAGAGCAGGACAAAAGGTGT
TTGACATCTATCTAAACAATGAGATTAAAAAGGAGAAATTTGATGTTTTGGCTGGAGGGT
CCAAGAACAGTTACACTGCCTTGAACATTTTCAGCAAATGGATCACTCAATATAACCTTAG
TCAAGGCATCTGGATCTGAGTTTGGACCCCTTTTGAATGCCTATGAAATCCTGCAGGCAC
GGTCGTGGATTGAAGAGACCAACCAAAAAGATTGTAAGTGTACAATAAGATGCTAAATTG
ANGAATTTTTATTAACTTTAAACCAATTTTCTATTTTCATTCTCCCTCTTACACTAACACTT
TTTTTCCTTTTCAGTGGAGTTATTCAGAAGATGAGAGAAGAACTGCTGCTGCACAACCAA
GAAAATGAAGCATTGGAGAGTTGGAGTGGAGACCCCTTGTATGATTTTTCCCTGGAAAGGA
ATAACATGTGATGATTCAACTGGTTCATCTATTATCACTAAGCTGTAAGTCCCTCCACTT
TTTAGGGTCTGTTTAGATTTGCTTATTTGAGTTTATCTATTGAAATAAACACTTATGACA
CTGTTTGGAAAGAGCTTATGAAAACAACTTATAGTTTATACGAAAAACAAGTTGACTTTGTT
ATATCTATTTTATAGAAATAGCTTATAAGTAAGAACTTATATGATAAGCGTTTATGCTAT
AAACGCTCAATTTAACTGTTTATCCAAACAGGACTTTAATGCATCCATATGTGTTTACAA
GTTTTGCTCTATTTTCTGCAGGGATCTTCTTCCAATAATCTCAAGGGAGCAATTCCTTC
CATTGTCCTAAGATGACCAATTTACAAATACTGTACTGCTTCTAAATCTCATTTAAAAA
CATCCACATTATTTTTGTGGGAACAGTGGACTGTTTTTATCCATTAGTTAATAATGTTGT
AACATTTTTGTTATGCGCAGGAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
TTCCACCGTCCCTCCTTGGCTGATATCATTTGTAATTATCTCTTCTCATGTTTAAACAAATGAA
CTAGGATGATACTAATATGAGCTTATAGCACCTTCTATTATCCTTGTAAATGTTAACATAA
ACAAGCACATGCAGATTATAGAACTAAAATATGATAGAATATGCTGTATATAGTCCCAAC
TCATGTTGGTTACATAAGGTAGATATTAGACAGTCTCAATGCTGAAACCAATATCTGTGT
GCATGATGCATCTTAATCTTTAAAAGATTTTATGATTAAGGCTGTTCTCTATGGCACTTC
AGTAGGAGCAGTAACACATCTATGTGAAAGTTCCATATGAAATCCTTGAGAAATATGTTT
TGACATTATGTTTCATATATTGCTGAATTTCTTCTTCCCAAGGTTTGAATCTTAGAGA
AATTTCTACAGTTATTAGTGCCATTGAAAAATTACTACTAAATCTTTTTCAAATGATTG
ATCTACATTTATGGATCAGGGATCTTAGCTACAATGATCTTTCAGGATGGCTTCCAGAAT
CCATTATCTCACTGCCACATTTAAAATCATTGTAAGTTTTATATGTTGGCATTCTACTTC
ATCCATATTAGGAAGCTATTTTCGATTGTTGTATATTTTAAATATACTTCATTTATTCAG
CTTGTATTTGATTTATTTCCAGATATTTTGGCTGCAATCCATCTATGAGTGACGAAGATA
CAACAAAGTTGAACAGTTCACTAATCAATACAGAGTATGAAGTATTATTGATGCATAACA
ATAGAAGTTTTTAAAAATAAAATAAAATAAATCAAAATTTTACTCGTTTTCTGTTTTACAG
TTATGGGAGATGCAAAGCAAAAAACCAAGTTTGGACAAGTATTCGTGATTGGAGCTAT
TACAAGTGGATCACTTTTGATTACTTTGGCTGTTGGAATCTATTTTTTGGCGTTATAG
ACACAAGTCAATTACTTTGGAAGGATTTGGAAAGACCTACCCAATGGCAACAAGTANGCA
TTTTATCTTTTATAATCTATTGTAACATATGTCTATCTGCATTCTAATGTAACATATTTG
CTCTTTAGACTAACAAACAAAACTATAGTCACATCATTCTGCATAGACAATCTCAGATC
ACAAGATCTAAGTGTTATCTTTAACATAGAGCATGGTTTACAAATACAGGGAATACATTT
TTATAACCCTGAGATTAAATCAACTTTTCTTATACAGTCAGACAGGATCTGCACATAATG
TTCAACACAATGATAAAATAAGAGAAATATATTTAAAAAACTAACATCCAGTGTGAATT
CCTAGATATATGGCTACAAAAAATCTATATTCCATCTGTACAATAACAAGAATATCAAT
TGAAAAAAGAATACTTATGAACATGATAGATGATATTAGTTTCTGTAGAGGATCATTACT
ATGATGGAAGTGATGATTTTAGCATGTTTGGTTTGGTCTAATCTACATTGTATATATGAC
ATTCTCCAACAAAAACATGCAGATATAATTTTCTCTTTGCCAAGCAAAGACGATTTCTTC
ATAAAGTCTGTATCAGTTAAACCGTTCACTTTGGAGTATATAGAGCAGGCTACAGAACAG
TACAAAACCTTTAATAGGTGAAGGAGGATTTGGCTCTGTTTACAGGGGCACTCTAGACGAT
GGTCAAGAAGTGGCAGTGAAAGTGCGGTCATCCACATCAACTCCGGGAACCCGAGAATTT
GATAATGAGGTATAATATGCATTATCACTTTATTAGAGCAACAAGATTTCCACAGAATGT
TCCTAATTAAAGCTTTGGCCATATATATAGTAGTTTAAAGAATTTGTTCCGGTTCCTATCAC
ACCATCTGATTGGTCTATGATGGATCATGCAGCTAAACCTACTTTTCAGCTATACAACATG
AGAACCTGGTGCCTCTTCTGGGTACTGTAATGAGTATGATCAACAAATTTCTCGTGTATC

11/60

Figure 8. (cont. 3/3)

CTTTCATGTCCAATGGCTCTTTGCTAGATAGACTATACGGTAAATATCTCACAATTCTTT
TAGCAGATATGTTTTAATACACAGTAATTCTGAAATTTAATTTTCATAACAGGGGAAGCAT
CAAAGAGAAAAATATTAGACTGGCCAACTAGACTCTCTATTGCTCTCGGTGCAGCTCGAG
GCAAGTACCATGATGTTGTTTTACTTTTACTAAGTTGTGCATATGTACTAATTAAGCTTG
CATCATATCAGAGAGGATAAAACATGAAATTGTATGTAAACATTTTCATATGAAAACATAA
AACTAATTCAAATGAACAATGAAAAAGAGAAGCGTATACAATGATTATGTCTTACCACA
GGCGCAGGGACTTGAATATAAGATACATATTGATTTCTCTACTGAACATATAAACTGGAG
ACAATCACAAGCAAAATGATGAGGCTATATTGATTGACAAGGACAATCATATTGTTTTAC
TGTGATTAGAACGATACAGTAAAGATATTTTAGAGGGAATACTGTTTACTACTATAAGAG
TGACTGATAGTTCCGATGTTGCTAATGGAAATTGATAAGATGGTCAACATTGACTTAGTG
AAATCTGTCTGACAATAAGTATTTTGGCCCTCACGGATCTACTTTGGCATAAAGAAAAA
CTTCTTGAAATGATGCATCCTAAAGTTCTTTATTCTTCACATAATAAACTCAGTTTTTGA
CCTAACTGACTTCACTACAGTTTTCGTCCATAACATATTTAGGACACTAAAATGATCATA
AGTTTTCTGATTGAGCAGGTTTGGCATATCTTCACACATTTCCAGGACGTCTGTAAATAC
ACAGGACGTAAAATCGAGCAATATACTGGATCAGAGCATGTGTGCTAAGTTGTGCAG
ATTTTGGTTTTCTCAAAATATGCTCCTCAGGAAGGAGACAGTTATGTTTCCCTTGAAGTAA
GAGGAAGTGCAGGGTATCTGGATCCTGAGTAAGTGAAGTTAAATCCGTCTTAATCTGAAT
GTCTCATACTGCTCTCTCATCTATCACTTTTCAGCAAAATATTCAATACATTTCCACACA
CAAGTTTGTGGAACCATCAGAAAAGTGAAGAAATGTTTCTGATTATATTCCTATTCTGA
ATCTGAAGTGTATATTAATTTTAAACATACAAAAGTTCATATGGTGACTACAA
ATGATAAGAGATATTTATCCATGTGAATTTGAATGGGATGAAGAAGAATCTAACATACCT
AGTCCCATTTGTTTACATGTTACAGAGAGTAACCATGACATGAACAAATTTCAACAATAG
CTGACGTTTATCCATATGCAGGTACTACAAAACCCAGCAATTATCTGAAAAAGTGATGT
TTTCAGCTTTGGTGTGGTTCTTCTTGAATTTGTAAGCGGACGGGAACCTCTCAACATAAA
GAGACCAAGGATCGAGTGGAGCTTGGTTGAATGGGTATGATCCATGACCCCATTTTTTTT
ACAAATATAATTTTCAAGTGGTTCATATATTTTAGTAGGTACTAACACATAAATTTTCATG
CAGGCTAAACCATACATAAGAGCATCAAAGGTGGATGAAATTTGTAGATCCTGGCATCAAG
GGAGGATATCATGCAGAGGCATTGTGGAGAGTTGTGGAAGTAGCACTACAATGTCTAGAA
CCCTACTCAACATATCGGCCATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTC
ATTATTGAAAACAATGCATCTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAAC
CGCTACTCAATTGTTATGGACAAACGGGCGCTGCCTTCACTACATCTACAGCAGAATCA
ACTATCACAACCCAAACCTTGTACACCCTCAACCGAGATAGTAAATGGGTGATGGAAT
TCTTTTGATTTGTTTTGATCATTGCTTTAGTAATATCACATTTTAAATGGTAAAGGAGA
AAAATACTACTTCTGATTGTATTTCCATCCACTCTATGTTTCTTGAAACTGAATCTCTCT
TGCTCAGCCCCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAAACCA
TATGGTGCATAATTTAAAAGCCATATCATATCATTGCGCAAGTCCAAAGTAAAAATTTCA
CAAAGTAGTTAGATTGCGATTTAGTCTATAGACACTTCAACAGAGCTATATACACTATGG
TTGACTTGCGACTAATTCGCTCAAGCAGGAGGAACACATATATATGGGAAACACTTGTAG
AACTATTTTGTATAGAAATGAAAATATTTCCGTTTTTAATTTTTGTTTTCATTAAGT
TTGAATTATTTTACTTTGGTTTATACATGATTCTAAATTATGTTTGAAGTTATGGTAAG
AAAGTTGTCATTGCGAAAAAATACGGCCGCAATTGCCGCTGTGCATCTTCAGCTCAAGTG
ATTGTCTCTTGGTGATTGTTCAACAGACCCTCAACATTGCTCATCCTATCAGAACCCAAC
TTCCAATAGGCATTTGTTCTACTTAAAGAAAATCGGAAGCTAGC

Figure
9.

A.

Forward primers:		Length (base)	Start point (bp)	Exon/Intron position
Pr_RKU5:	5'-AGGGGTTACTTGCTCTTTTGTGGT-3'	24	72	prom.
Pr_RKU6	5'-ATCTCTAACTCGTCTCCATCTC-3'	23	847	prom.
Pr_RKU4	5'-TTCCAGGCCTAAAGTCAAACACCA-3'	24	897	E 1
Pr_RK-ATG	5'-TCTCTTTGGTACCATGATGGAGTTACAAGTTA-3'	32	1139	E 1
Pr_RKE2U	5'-ATGGATCCGGGTTTGAGAGCATAGCATG-3'	28	1563	E 2
Pr_RKU3	5'-TGAGGGAGTTTTAGAGCCACCAAG-3'	26	1892	E 2
Pr_RKU2	5'-CACCTGAGCGATTGGAGTTC-3'	21	2946	E 3
Pr_Pst-A5	5'-GAAATCCTGCAGGCACGGTCGTGG-3'	24	3222	E 3
Pr_RKU7	5'-AAGCATTGGAGAGTTGGAGTGG-3'	22	3426	E 4
Pr_RKU2X	5'-AAACAAGCACATGCAGATTAT-3'	21	4138	I 6
Pr_RK-ECO-5'	5'-GGCTGTTGGAATTCTATTTTTTGC-3'	25	4947	E 9
Pr_RKU1	5'-TACAGGGGCACTCTAGACGATGGT-3'	24	5679	E 10
Pr_Pst-B5	5'-AAGAGGAACTGCAGGGTATCTGG-3'	23	6956	E 13
Pr_RKU0	5'-GCAGAGGCATTGTGGAGAGT-3'	20	7631	E 15

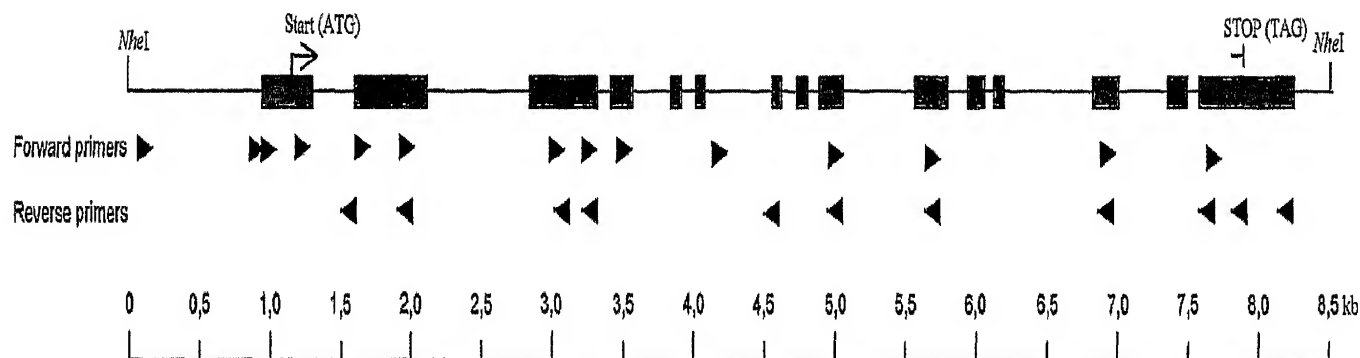
Reverse primers:

Pr_RKD5	5'-AGGGTGGTTTTGGATCTGTGTAA-3'	24	1621	E 2
Pr_RKD4	5'-TTAAACACTAGTACCGAAACC-3'	22	2028	E 2
Pr_RK_D3	5'-ACCCTCCAGCCAAAACATCAAA-3'	22	3120	E 3
Pr_Pst-A3	5'-ACGACCGTGCCTGCAGGATTTC-3'	23	3246	E 3
Pr_RKD2x	5'-TAAATGTGGCAGTGAGATAATGGA-3'	24	4582	E 7
Pr_RK-ECO-3'	5'-GCAAAAAAATAGAATTCCAACAGCC-3'	25	4972	E 9
Pr_RKD2	5'-CCCTGAGTTGATGTGGATGAC-3'	21	5746	E 10
Pr_RKD7	5'-GATACCCTGCAGTTCCTCTTA-3'	21	6976	E 13
Pr_RKD1	5'-GGCCGATATGTTGAGTAGGGTTCT-3'	24	7698	E 15
Pr_RK-STOP	5'-AAAAGAGCTCCATCGACCCATTTACTATC-3'	29	7924	E 15
Pr_RKD0	5'-GAATTAGTCGCAAGTCAACC-3'	20	8236	E 15

■ = exon

12/60

B.



13/60

Figure 10.

TTCCAGGCCTAAAGTCAAACACCATATTTTAAACAATATTCTTTCTTCTACAGCTTTTATA
CAAGTTCACTATATTATAGGATTGATCAGGGTTCATTTTTCTTTCTTTGAAAAATCTCT
AAGGGGTGCTGTTTCCAAGGCAGAGAATGAAATAGAATTCAGAAAGAAATTTTATGTTACT
ATAAAGGAAAGATGAAAAGTTAGTTAGCATGGATTCAAGTTTGATAACCCCTATGGGGTAA
AATCTCTTTCAGATTATGATGGAGTTACAAGTTATTAGGATATTTAGATTGGTTGTGGCA
TTTGTTCTTTGTTGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAGGGTTTGAG
AGCATAGCATGTTGTGCTGATTCAAATTACACAGATCCAAAAACCACCTAACTTATACA
ACAGATCACATCTGGTTCTCTGATAAAAAGAAGTTGCAGACAAATACCCGAAATTTTGTTT
AGCCACAGAAGCAATAAAAAATGTTCGAAAATTTGAAATATATGAAGGAAAGAGATGTTAT
AATTTGCCAACAGTTAAGGATCAAGTATATTTGATAAGGGGCATATTTCCCTTTTGATAGT
TTAAATTCTTCGTTTATGTTTCGATCGGGGTAACAGAACTAGGCGAATTAAGATCGTCT
AGGCTCGAGGACTTGGAATTGAGGGAGTTTTTTAGAGCCACCAAAAGACTACATAGATTTT
TGCTTATTGAAGGAGGATGTCAATCCCTTCATTTCTCAGATTGAATTGAGGCCATTACCT
GAAGAATACCTACATGGTTTCGGTACTAGTGTTTTAAACTGATAAGCAGAAACAATCTT
GGTGACACAAATGATGATATAAGGTTCCAGATGACCAAAATGATAGAATCTGGAAACGG
AAAGAACTTCAACTCCAACATCTGCCCTTCCACTGTCTTCAATGTCAGCAATGTTGAC
CTCAAAGACAGTGTCACACCTCCTCTACAAGTCTTACAAACAGCTCTTACTCaCCCTGAG
CGATTGGAGTTCGTCCATGATGGCCTCGAGACCGATGATTATGAATACTCTGTGTTTCTC
CACTTTCTTGAACATAAATGGCACTGTGAGAGCAGGACAAAGGGTGTGACATCTATCTA
AACAATGAGATTAAAAGGAGAAATTTGATGTTTTGGCTGGAGGGTCCAAGAACAGTTAC
ACTGCCTTGAACATTTTCAGCAAATGGATCACTCAATATAACCTTAGTCAAGGCATCTGGA
TCTGAGTTTGGAACCTTTTGAATGCCTATGAAATCCTGCAGGCACGGTCTGGATTGAA
GAGACCAACCAAAAAGATTTGGAAGTTATTGAGAAGATGAGAGAAGAACTGCTGCTGCAC
AACCAAGAAAATGAAGCATTGGAGAGTTGGAGTGGAGACCTTGTATGATTTTCCCCTGG
AAAGGAATAACATGTGATGATTCAACTGGTTTCATCTATTATCACTAAGCTGGATCTTTCT
TCCAATAATCTCAAGGGAGCAATTCCTTCCATTGTCACTAAGATGACCAATTTACAAATA
CTGAACCTGAGCCACAACCAAGTTCGATATGTTATTCCCCTCGTTTCCACCGTCTCCTTG
CTGATATCATTGGATCTTAGCTACAATGATCTTTTCAGGATGGCTTCCAGAATCCATTATC
TCACTGCCACATTTAAAATCATTATATTTTGGCTGCAATCCATCTATGAGTGACGAAGAT
ACAACAAAGTTGAACAGTTCACTAATCAATACAGATTATGGGAGATGCAAAGCAAAAAA
CCAAAGTTTGGAACAAGTATTCGTGATTGGAGCTATTACAAGTGGATCACTTTTGATTACT
TTGGCTGTTGGAATTTCTATTTTTTTTGGCGTTATAGACACAAGTCAATTACTTTGGAAGGA
TTTGGAAGACCTACCCAATGGCAACAAATATAATTTTCTCTTTGCCAAGCAAAGACGAT
TTCTTCATAAAGTCTGTATCAGTTAAACCGTTCACTTTGGAGTATATAGAGCAGGCTACA
GAACAGTACAAAACCTTAAATAGGTGAAGGAGGATTTGGCTCTGTTTACAGGGGCACTCTA
GACGATGGTCAAGAAGTGGCAGTGAAAGTGCGGTCACTCCACATCAACTCAGGGAACCCGA
GAATTTGATAATGAGCTAAACCTACTTTTCAGCTATACAACATGAGAACCTGGTGCCTCTT
CTGGGTTACTGTAATGAGTATGATCAACAAATTTCTCGTGTATCCTTTTCATGTCCAATGGC
TCTTTGCTAGATAGACTATACGGGGAAGCATCAAAGAGAAAAATATTAGACTGGCCAACCT
AGACTCTCTATTGCTCTCGGTGCAGCTCGAGGTTTGGCATATCTTTCACACATTTCCAGGA
CGTTCTGTAATACACAGGGACGTAAAATCGAGCAATATACTGCTGGATCAGAGCATGTGT
GCTAAGGTTGCAGATTTTGGTTTCTCAAAATATGCTCCTCAGGAAGGAGACAGTTATGTT
TCCCTTGAAGTAAGAGGAACGTCAGGGTATCTGGATCCTGAGTACTACAAAACCCAGCAA
TTATCTGAAAAAAGTGATGTTTTTCAGCTTTGGTGTGGTTCTTCTTGAAATTTGTAAGCGGA
CGGGAACCTCTCAACATAAAGAGACCAAGGATCGAGTGGAGCTTGGTTGAATGGGCTAAA
CCATACATAAGAGCATCAAAGGTGGATGAAATTGTAGATCCTGGCATCAAGGGAGGATAT
CATGCAGAGGCATTGTGGAGAGTTGTGGAAGTAGCACTACAATGTCTAGAACCTTACTCA
ACATATCGGCCATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTCATTTATGAA
AACAATGCATCTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAACCGCTACTCA
ATTGTTATGGACAAAACGGGCGCTGCCTTCACTACATCTACAGCAGAATCAACTATCACA
ACCCAAACCTTGTACACCCCTCAACCGAGATAGTAAATGGGTGATGGAATTTCTTTTGAT
TTGTTTTTGTATTTGCTTTAGTAATATCACATTTTAAATGGTAAAGGAGAAAAATACTA
CTTCTGATTGTATTTCCATCCACTCTATGTTTCTTGAAACTGAATCTCTCTTGCTCAGCC
CCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAAACCATATGGTGCA
TAATTTAAAGCCATATCATATCATTTGCCAAGTCAAAGTAAAAATTTACAAACTAGT
TAGATTGCGATTTAGTCTATAGACACTTCAACAGAGCTATATACACTATGGTTGACTTGC
GA

14/60

Figure 11. (1/5)

NheI

GCTAGCTTAATTTGTGAAGTAGTTTTTCAGTCAGCTCATGCTGAACAGTTATGGTAGTTTT	BAC A17
ACAAGGATGTAAGGGGTTACTTGCTCTTTTGTGGTATCATTGAGCCACTTTCCACTTTCT	BAC A17
TAATTCTTTGCTATCAAATTTAAATATATTATATTATTGTAGGAAATATGAAAAAGAAAC	BAC A17
CCTAKATGCTTTAATTGTGCTAAGCATATAAGATACAAAAAAGTCTAATCTCTT	BAC A17
TTGATAGGAACTGACACAAGGTTTCGATCCTCAGAACCACCTATGAGACTAATTTTTGCT	BAC A17
TGAATTGGTTAATCAACTGCATCGCACTCATTGAATAAAAAAGTACGTTTAGATTGACGG	BAC A17
TGAATTTAACATAATCACAGTTGACCACTTTGATTTTTCACGAAAGTTGCATGTAACTTTC	BAC A17
GTGCTTGAATATAAAGTCACGGTGCCACAATGGTTTTGTCAATATCGCCGKAATCCAAA	BAC A17
CATGCACTTAATATGTTTTTATATTAATGGAACCTATGTATATAAATATAAAATACTTA	BAC A17
AGCATATATATCCTTGCAACAATACCAAAAAAGTTTAAATTTGCTCCAAAGATACTATT	BAC A17
CACGGTTTTCTATTGTGATGCATCCATGGAATGAGAAGGAAGAAACAGAAGAAGACGAAG	BAC A17
GACTGGTCTGTCCAATTATTTAAACCATATGAATATTTTATATACAGATAATGAAGCAA	BAC A17
CCACCACTAAACAATAAGAATAAATAAATGTTCAACATCATCAATAATAAGAATAAT	BAC A17
AATGTTTAAACATCAACAATAAACTATAGGAAAAAACATAATCAACTATGCATTGTACT	BAC A17
AATTCATCTCTAACTCGTCTTCCATCTCTTTCCTAGCTACCTCCTGCAGTTTCT	BAC A17
TTCCAGGCCTAAAGTCAAACACCATATTTTAACAATATTCTTTCTTCTACAGCTTTTATA	BAC A17
TTCCAGGCCTAAAGTCAAACACCATATTTTAACAATATTCTTTCTTCTACAGCTTTTATA	cDNS A17
CAAGTTCACATATATTATAGGATTGATCAGGGTTCATTTTTCTTTCTTTTGAAAAATCTCT	BAC A17
CAAGTTCACATATATTATAGGATTGATCAGGGTTCATTTTTCTTTCTTTTGAAAAATCTCT	cDNS A17
AAGGGGTGCTGTTTCCAAGGCAGAGAATGAAATAGAATTCAGAAGAATTTTTATGTTACT	BAC A17
AAGGGGTGCTGTTTCCAAGGCAGAGAATGAAATAGAATTCAGAAGAATTTTTATGTTACT	cDNS A17
ATAAAGGAAAGATGAAAAGTTAGTTAGCATGGATTCAAGTTTGATAACCCATATGGGGTAA	BAC A17
ATAAAGGAAAGATGAAAAGTTAGTTAGCATGGATTCAAGTTTGATAACCCATATGGGGTAA	cDNS A17
AATCTCTTTTCAAGATTATGATGGAGTTACAAGTTATTAGGATATTTAGATTGGTTGTGGCA	BAC A17
AATCTCTTTTCAAGATTATGATGGAGTTACAAGTTATTAGGATATTTAGATTGGTTGTGGCA	cDNS A17
TTTGTTCTTTGTTTGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAGGTTAGTAG	BAC A17
TTTGTTCTTTGTTTGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAG	cDNS A17
CTAAATACTATAATTCTTTAAGATCATAATAATATCTATTACTTGATTCTTTTCCCTTTA	
AACATAGAAAACAACATATTTTAATTAACATGAAAGGCCATGGGATGATCATAATTAATA	
ATGAGAAGAAAATAAATAAGAACTTGTTTTTTTTTACAAGGATTGTAATAAGAACTAGT	I1
AGTTTCATACTTTCAATACTGAGAATCTTGAAACAATTTCACTTTTCTTTATGTTGCTA	
GAATTTCTTTCAAGGGAATCCAAATTTGTACAAAATGAATTTAACTTGTGACATTTTC	
CTTGTA <u>G</u>	
GGTTTGAGAGCATAGCATGTTGTGCTGATTCAAATTACACAGATCCAAAAACCACCTAA	BAC A17
GGTTTGAGAGCATAGCATGTTGTGCTGATTCAAATTACACAGATCCAAAAACCACCTAA	cDNS A17
CTTATACAACAGATCACATCTGGTTCTCTGATAAAAGAAGTTGCAGACAAATACCCGAAA	BAC A17
CTTATACAACAGATCACATCTGGTTCTCTGATAAAAGAAGTTGCAGACAAATACCCGAAA	cDNS A17
TTTTGTTTAGCCACAGAAGCAATAAAAAATGTTGAAAAATTTGAAATATATGAAGGAAAGA	BAC A17
TTTTGTTTAGCCACAGAAGCAATAAAAAATGTTGAAAAATTTGAAATATATGAAGGAAAGA	cDNS A17
GATGTTATAATTTGCCAACAGTTAAGGATCAAGTATATTTGATAAGGGGCATATTTCCCT	BAC A17
GATGTTATAATTTGCCAACAGTTAAGGATCAAGTATATTTGATAAGGGGCATATTTCCCT	cDNS A17
TTGATAGTTTAAATTCTTCGTTTTATGTTTCGATCGGGGTAACAGAACTAGGCGAATTAA	BAC A17
TTGATAGTTTAAATTCTTCGTTTTATGTTTCGATCGGGGTAACAGAACTAGGCGAATTAA	cDNS A17
GATCGTCTAGGCTCGAGGACTTGGAATTTGAGGGAGTTTTTAGAGCCACCAAAGACTACA	BAC A17
GATCGTCTAGGCTCGAGGACTTGGAATTTGAGGGAGTTTTTAGAGCCACCAAAGACTACA	cDNS A17
TAGATTTCTGCTTATTGAAGGAGGATGTCAATCCCTTCATTTCTCAGATTGAATTGAGGC	BAC A17
TAGATTTCTGCTTATTGAAGGAGGATGTCAATCCCTTCATTTCTCAGATTGAATTGAGGC	cDNS A17

15/60

Figure 11. (cont. 2/5)

CATTACCTGAAGAATACCTACATGGTTTCGGTACTAGTGTTTTAAAACTGATAAGCAGAA	BAC A17
CATTACCTGAAGAATACCTACATGGTTTCGGTACTAGTGTTTTAAAACTGATAAGCAGAA	cDNS A17
ACAATCTTGGTGACACAAATGATGATATAAGGATATGTGATCTTACTTTATTTTTAGGTAG	BAC A17
ACAATCTTGGTGACACAAATGATGATATAAG	cDNS A17
ATTCCACCTCTATTTTACAGGGAGTGCTCTCAGGAAACCTAAAAGGCTTAGGGTTTGTCTC ACTACTTGGTTCTTATGGAAGCATCATATGTTCTACTTAGTAGATATATATGTATATAAA AATCAACATCTTTCTCCGTAAGGGATAGAAATTGCAATTCATGTTATGTTGAGCATATTT AAACCAAATAATTTGAGGAATATGATGCATGAACCTTTCTGCCAAATGCATGACATAAC CTATGTTTCACTTTACATATAGAATTATAAGATGTGTTTACATCTTTATATTAACCTTGA TCTTTATGATGCACTGAGATACACCTGAATTTAAATAACAAATGGAGAAGCAAGAATCTA GAACTCTACCATTCAATCCATTGTAATATGGAACCTCACAAGAATTAACCTATTGCCACTTC AGATTAAATACCCTAAGAGTGTTTTTGTGGATGAGGGAATGTTTTGAGGGAATGTAATGT TTTGAGGGAATTCCTCAAAACATTACATTACCTCAAAATATTCCCCATCCAAACACACT ATAAAGGATCCTAATCTAGTGAGCCTTAGTTTTCTATTGAAGTGATTTGTGTTTAGAGT TTGGTCTGCATCTGCAACTTTGTTTTGTTGCTGTAATGCCATGACTATGAACACTGACAG AAAGGTCTCATAATATCGGTATCTATCAATTTTAG	I2
GTTCCCGATGACCAAAATGATAGAATCTGGAAACGGAAGAACTTCAACTCCAACATC	BAC A17
GTTCCCGATGACCAAAATGATAGAATCTGGAAACGGAAGAACTTCAACTCCAACATC	cDNS A17
TGCCCTTCCACTGTCTTTCAATGTCAGCAATGTTGACCTCAAAGACAGTGTCACACCTCC	BAC A17
TGCCCTTCCACTGTCTTTCAATGTCAGCAATGTTGACCTCAAAGACAGTGTCACACCTCC	cDNS A17
TCTACAAGTCTTACAAACAGCTCTTACTCACCTGAGCGATTGGAGTTCGTCCATGATGG	BAC A17
TCTACAAGTCTTACAAACAGCTCTTACTCACCTGAGCGATTGGAGTTCGTCCATGATGG	cDNS A17
CCTCGAGACCGATGATTATGAATACTCTGTGTTTCTCCACTTTCTTGAACATAATGGCAC	BAC A17
CCTCGAGACCGATGATTATGAATACTCTGTGTTTCTCCACTTTCTTGAACATAATGGCAC	cDNS A17
TGTCAGAGCAGGACAAAGGGTGTTTGACATCTATCTAAACAATGAGATTAAAAAGGAGAA	BAC A17
TGTCAGAGCAGGACAAAGGGTGTTTGACATCTATCTAAACAATGAGATTAAAAAGGAGAA	cDNS A17
ATTTGATGTTTTGGCTGGAGGGTCCAAGAACAGTTACACTGCCTTGAACATTTTACGCAAA	BAC A17
ATTTGATGTTTTGGCTGGAGGGTCCAAGAACAGTTACACTGCCTTGAACATTTTACGCAAA	cDNS A17
TGGATCACTCAATATAACCTTAGTCAAGGCATCTGGATCTGAGTTTGGACCCCTTTTGAA	BAC A17
TGGATCACTCAATATAACCTTAGTCAAGGCATCTGGATCTGAGTTTGGACCCCTTTTGAA	cDNS A17
TGCCTATGAAATCCTGCAGGCACGGTCGTGGATTGAAGAGACCAACCAAAAAGATTGTAA	BAC A17
TGCCTATGAAATCCTGCAGGCACGGTCGTGGATTGAAGAGACCAACCAAAAAGATT	cDNS A17
GTGTACAATAAGATTGCTAAATNGATGAATTTTATTAACCTTTAACCAATTTTCTATTT CATTTCTCCCTCTTACACTAACACTTTTTTTCCTTTTCA	I3
TGGAAGTTATTTCAGAAGATGAGAGAAGAACTGCTGCTGCACAACCAAGAAAATGAAGCAT	BAC A17
TGGAAGTTATTTCAGAAGATGAGAGAAGAACTGCTGCTGCACAACCAAGAAAATGAAGCAT	cDNS A17
TGGAGAGTTGGAGTGGAGACCCCTTGATGATTTTCCCTGGAAAGGAATAACATGTGATG	BAC A17
TGGAGAGTTGGAGTGGAGACCCCTTGATGATTTTCCCTGGAAAGGAATAACATGTGATG	cDNS A17
ATTCAACTGGTTCATCTATTATCACTAAGCTGTAAAGTCCTTCCACTTTTTAGGGTCTGTT	BAC A17
ATTCAACTGGTTCATCTATTATCACTAAGCT	cDNS A17

16/60

Figure 11. (cont. 3/5)

TAGATTTGCTTATTTGAGTTTATCTATTGAAATAAACACTTATGACACTGTTTGGAGAG CTTATGAAACAACCTTATAGTTTATACGAAAACAAGTTGACTTTGTTATATCTATTTTAT AGAAATAGCTTATAAGTAAGAAGCTTATATGATAAGCGTTTATGCTATAAACGCTCAATTT AACTGTTTATCCAAACAGGACTTTAATGCATCCATATGTGTTTACAAGTTTGTCTCTATT TTCTGCAG	I 4
GGATCTTTCTTCCAATAATCTCAAGGGAGCAATTCCTTCCATTGTCACATAAGATGACCAA GGATCTTTCTTCCAATAATCTCAAGGGAGCAATTCCTTCCATTGTCACATAAGATGACCAA	BAC A17 cDNS A17
TTTACAAATACTGTACTGCTTCTAAATCTCATTTAAAAACATCCACATTATTTTTGTGGG TTTACAAATACT	BAC A17 cDNS A17
AACAGTGGACTGTTTTTATCCATTAGTTAATAATGTTGTAACATTTTTGTTATGCGCAG	I 5
GAACCTGAGCCACAACCAGTTTCGATATGTTATTCCCTCGTTTCCACCGTCCTCCTTGCT GAACCTGAGCCACAACCAGTTTCGATATGTTATTCCCTCGTTTCCACCGTCCTCCTTGCT	BAC A17 cDNS A17
GATATCATTGTAATTATCTCTTCTCATGTTTAAACAAATGAACTAGGATGATACTAATATG GATATCATT	BAC A17 cDNS A17
AGCTTATAGCACTTCTATTATCCTTGTAATGTTAACATAAACAGCACATGCAGATTAT AGAATAAAATATGATAGAATATGCTGTATATAGTCCCAACTCATGTTGGTTACATAAGG TAGATATTAGACAGTCTCAATGCTGAAACCAATATCTGTGTGCATGATGCATCTTAATCT TTAAAAGATTTTATGATTAAAGGCTGTTCTCTATGGCACTTCAGTAGGAGCAGTAACACAT CTATGTGAAAGTTCCATATGAAATCCTTGAGAAATATGTTTGTGACATTATGTTTCATATA TTGCTGAATTTCTTTCTTCACCAAGGTTTGAATCTTAGAGAAATTTCTACAGTTATTAG TGCCATTGAAAAATTACTACTAAATCTTTTCAAAATGATTGATCTACATTTATGGATCAG	I 6
GGATCTTAGCTACAATGATCTTTTCAGGATGGCTTCCAGAATCCATTATCTCACTGCCACA GGATCTTAGCTACAATGATCTTTTCAGGATGGCTTCCAGAATCCATTATCTCACTGCCACA	BAC A17 cDNS A17
TTTAAATCATTGTAAGTTTATATGTTGGCATTTCTACTTCATCCATATTAGGAAGCTAT TTTAAATCATT	BAC A17 cDNS A17
TTTCGATTGTTGTATATTTTTAATATACTTCATTTATTTCAGCTTGTATTTGATTTATTTT CAG	I 7
ATATTTTGGCTGCAATCCATCTATGAGTGACGAAGATACAACAAAGTTGAACAGTTCACT ATATTTTGGCTGCAATCCATCTATGAGTGACGAAGATACAACAAAGTTGAACAGTTCACT	BAC A17 cDNS A17
AATCAATACAGAGTATGAAGTATTATTGATGCATAACAATAGAAGTTTTTAAATAAAAT AATCAATACAGA	BAC A17 cDNS A17
AAAATAAATCAAAATTTTACTCGTTTCTGTTTTCACAG	I 8
TTATGGGAGATGCAAAGCAAAAAAACAAAGTTTGGACAAGTATTCGTGATTGGAGCTAT TTATGGGAGATGCAAAGCAAAAAAACAAAGTTTGGACAAGTATTCGTGATTGGAGCTAT	BAC A17 cDNS A17
TACAAGTGGATCACTTTTGATTACTTTGGCTGTTGGAATTCTATTTTTTTGCCGTTATAG TACAAGTGGATCACTTTTGATTACTTTGGCTGTTGGAATTCTATTTTTTTGCCGTTATAG	BAC A17 cDNS A17
ACACAAGTCAATTACTTTGGAAGGATTGGAAGACCTACCCAATGGCAACAAGTANGCA ACACAAGTCAATTACTTTGGAAGGATTGGAAGACCTACCCAATGGCAACAA	BAC A17 cDNS A17
TTTTATCTTTTATAATCTATTGTAACATATGCTCTATCTGCATTCTAATGTAACATATTTG CTCTTTAGACTAACAAACAAAACTATAGTCACATCATTCTGCATAGACAATCTCAGATC ACAAGATCTAAGTGTTATCTTTAACATAGAGCATGGTTTACAAATACAGGGAATACATTT TTATAACCCTGAGATTAAATCACTTTTCTTATACAGTCAGACAGGATCTGCACATAATG TTCAACACAATGATAAAATAAGAGAAATATATTTAAAAAACTAACATCCAGTGTGAATT CCTAGATATATGGCTACAAAAAATCTATATTCATCTGTACAATAACAAGPATATCAAT TGAAAAAGAATACTTATGAACATGATAGATGATATTAGTTTCTGTAGAGGATCATTACT	I 9

17/60

Figure 11. (cont. 4/5)

ATGATGGAAGTGATGATTTTAGCATGTTTGGTTTGGTCTAATCTACATTGTATATATGAC
 ATTCTCCAACAAAACATGCAG

ATATAATTTTCTCTTTGCCAAGCAAAGACGATTTCTTCATAAAGTCTGTATCAGTTAAAC BAC A17
 ATATAATTTTCTCTTTGCCAAGCAAAGACGATTTCTTCATAAAGTCTGTATCAGTTAAAC cDNS A17

CGTTCACCTTTGGAGTATATAGAGCAGGCTACAGAACAGTACAAAACCTTTAATAGGTGAAG BAC A17
 CGTTCACCTTTGGAGTATATAGAGCAGGCTACAGAACAGTACAAAACCTTTAATAGGTGAAG cDNS A17

GAGGATTTGGCTCTGTTTACAGGGGCACTCTAGACGATGGTCAAGAAGTGGCAGTGAAAG BAC A17
 GAGGATTTGGCTCTGTTTACAGGGGCACTCTAGACGATGGTCAAGAAGTGGCAGTGAAAG cDNS A17

TGCGGTCATCCACATCAACTCAGGGAACCCGAGAATTTGATAATGAGGTATAATATGCAT BAC A17
 TGCGGTCATCCACATCAACTCAGGGAACCCGAGAATTTGATAATGAG cDNS A17

TATCACTTTATTAGAGCAACAAGATTTCCACAGAATGTTCTTAATTAAAGCTTTGGCCAT I10
 ATATATAGTAGTTTAAGAATTTGTTCCGGTTCCTATCACACCATCTGATTGGTCTATGATG
 GATCATGCAG

CTAAACCTACTTTTCAGCTATACAACATGAGAACCTGGTGCCTCTTCTGGGTTACTGTAAT BAC A17
 CTAAACCTACTTTTCAGCTATACAACATGAGAACCTGGTGCCTCTTCTGGGTTACTGTAAT cDNS A17

GAGTATGATCAACAAATTCTCGTGTATCCTTTCATGTCCAATGGCTCTTTGCTAGATAGA BAC A17
 GAGTATGATCAACAAATTCTCGTGTATCCTTTCATGTCCAATGGCTCTTTGCTAGATAGA cDNS A17

CTATACGGTAAATATCTCACAATTCCTTTTAGCAGATATGTTTTAATACACAGTAATTCTG BAC A17
 CTATACG cDNS A17

AAATTTAATTTCATAACAG I11

GGGAAGCATCAAAGAGAAAAATATTAGACTGGCCAACTAGACTCTCTATTGCTCTCGGTG BAC A17
 GGGAAGCATCAAAGAGAAAAATATTAGACTGGCCAACTAGACTCTCTATTGCTCTCGGTG cDNS A17

CAGCTCGAGGCAAGTACCATGATGTTGTTTACTTTTACTAAGTTGTGCATATGTACTAA BAC A17
 CAGCTCGAG cDNS A17

TTAAGCTTGCATCATATCAGAGAGGATAAAACATGAAATTGTATGTAAACATTTTCATATG
 AAAACATAAACTAATTCAAATGAACAATGAAAAAGAGAAGCGTATACAATGATTATGT
 CTTACCACAGGCGCAGGGACTTGAATATAAGATACATATTGATTTCTCTACTGAACATAT
 AAACCTGGAGACAATCACAAGCAAATGATGAGGCTATATTGATTGACAAGGACAATCATA
 TTGTTTTACTGTGATTAGAACGATACAGTAAAGATATTTTAGAGGGAATACTGTTTACTA I12
 CTATAAGAGTGACTGATAGTTCGGATGTTGCTAATGGAAATTGATAAGATGGTCAACATT
 GACTTAGTGAAATCTGTCTGACAACTAAGTATTTTGGCCCTCACGGATCTACTTTGGCAT
 AAAGAAAAACTTCTTGAAATGATGCATCCTAAAGTTCTTTATTTCTCACATAATAAACTC
 AGTTTTTIGACCTAACTGACTTCACTACAGTTTTCTGTCATAACATATTTAGGACACTAAA
 ATGATCATAAGTTTTCTGATTGAGCAG

GTTTGGCATATCTTCACACATTTCCAGGACGTTCTGTAATACACAGGGACGTAAAATCGA BAC A17
 GTTTGGCATATCTTCACACATTTCCAGGACGTTCTGTAATACACAGGGACGTAAAATCGA cDNS A17

GCAATATACTGCTGGATCAGAGCATGTGTGCTAAGGTTGCAGATTTTGGTTTCTCAAAAT BAC A17
 GCAATATACTGCTGGATCAGAGCATGTGTGCTAAGGTTGCAGATTTTGGTTTCTCAAAAT cDNS A17

ATGCTCCTCAGGAAGGAGACAGTTATGTTTCCCTTGAAGTAAGAGGAACTGCAGGGTATC BAC A17
 ATGCTCCTCAGGAAGGAGACAGTTATGTTTCCCTTGAAGTAAGAGGAACTGCAGGGTATC cDNS A17

TGGATCCTGAGTAAGTGAACCTTAAATCCGTCTTAATCTGAATGTCTCATACTGCTCTCTC BAC A17
 TGGATCCTGA cDNS A17

ATCTATCACTTTTCCAGCAAAATATTCAATACATTTCCACACACAAGTTTGTGGAAACCAT
 CAGAAAACCTGAAGAAATGTTTCTGATTATATTCCTATTTCTGAATCTGAAGTGTATATTAA

18/60

Figure 11. (cont. 5/5)

TTTTAAACATACAAAACCTAAAGTTCCCATATGGTGACTACAAATGATAAGAGATATTTAT CCATGTGAATTTGAATGGGATGAAGAAGTCTAACATACCTAGTCCCATTGTTTCACA TGTTACAGAGAGTAACCATGACATGAACAAATTTACAATAGCTGACGTTTATCCATATG <u>CAG</u>	I13
GTACTACAAAACCCAGCAATTATCTGAAAAAGTGATGTTTTTCAGCTTTGGTGTGGTTCT GTACTACAAAACCCAGCAATTATCTGAAAAAGTGATGTTTTTCAGCTTTGGTGTGGTTCT	BAC A17 cDNS A17
TCTTGAAATTGTAAGCGGACGGGAACCTCTCAACATAAAGAGACCAAGGATCGAGTGGAG TCTTGAAATTGTAAGCGGACGGGAACCTCTCAACATAAAGAGACCAAGGATCGAGTGGAG	BAC A17 cDNS A17
CTTGGTTGAATGGGATGATCCATGACCCCATTTTTTTTACAAATATAATTTTCAAGTGG CTTGGTTGAATGG	BAC A17 cDNS A17
TTCATATATTTTAGTAGGTACTAACACATAACTTTTCATGC <u>CAG</u>	I14
GCTAAACCATAACATAAGAGCATCAAAGGTGGATGAAATTGTAGATCCTGGCATCAAGGGA GCTAAACCATAACATAAGAGCATCAAAGGTGGATGAAATTGTAGATCCTGGCATCAAGGGA	BAC A17 cDNS A17
GGATATCATGCAGAGGCATTGTGGAGAGTTGTGGAAGTAGCACTACAATGTCTAGAACCC GGATATCATGCAGAGGCATTGTGGAGAGTTGTGGAAGTAGCACTACAATGTCTAGAACCC	BAC A17 cDNS A17
TACTCAACATATCGGCCATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTCATT TACTCAACATATCGGCCATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTCATT	BAC A17 cDNS A17
ATTGAAAACAATGCATCTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAACCGC ATTGAAAACAATGCATCTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAACCGC	BAC A17 cDNS A17
TACTCAATTGTTATGGACAAACGGGCGCTGCCTTCAACTACATCTACAGCAGAATCAACT TACTCAATTGTTATGGACAAACGGGCGCTGCCTTCAACTACATCTACAGCAGAATCAACT	BAC A17 cDNS A17
ATCACAACCCAAACCTTGTACACCCCTCAACCGAGATAGTAAATGGGTCGATGGAATTCT ATCACAACCCAAACCTTGTACACCCCTCAACCGAGATAGTAAATGGGTCGATGGAATTCT	BAC A17 cDNS A17
TTTGATTTGTTTTTGATCATTGCTTTAGTAATATCACATTTTAAATGGTAAAGGAGAAAA TTTGATTTGTTTTTGATCATTGCTTTAGTAATATCACATTTTAAATGGTAAAGGAGAAAA	BAC A17 cDNS A17
ATACTACTTCTGATTGTATTTCCATCCACTCTATGTTTCTTGAACTGAATCTCTCTTGC ATACTACTTCTGATTGTATTTCCATCCACTCTATGTTTCTTGAACTGAATCTCTCTTGC	BAC A17 cDNS A17
TCAGCCCCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAAACCATAT TCAGCCCCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAAACCATAT	BAC A17 cDNS A17
GGTGCATAATTTAAAAGCCATATCATATCATTGCGCAAGTCCAAAGTAAAAATTTACAA GGTGCATAATTTAAAAGCCATATCATATCATTGCGCAAGTCCAAAGTAAAAATTTACAA	BAC A17 cDNS A17
ACTAGTTAGATTGCGATTTAGTCTATAGACACTTCAACAGAGCTATATACACTATGGTTG ACTAGTTAGATTGCGATTTAGTCTATAGACACTTCAACAGAGCTATATACACTATGGTTG	BAC A17 cDNS A17
ACTTGCGACTAATTCGCTCAAGCAGGAGGAACACATATATATGGGAAACACTTGTAGAAC ACTTGCGA	BAC A17 cDNS A17
TATTTTGTTTATAGAAATGAAAATATTTTCCGTTTTTAAATTTTGTTTTTCATTAAGTTTG AATTATTTTACTTTGGTTCATACATGATTCTAAATTATGTTTGAAGTTATGGTAAGAAA GTTGTCAATTGCGAAAAAATACGGCCGCAATTGCCGCTGTGCATCTTCAGCTCAAGTGATT GTCTCTTGGTGATTGTTCAACAGACCCCTCAaCATTGCTCATCTATCAGAACCCAACTTC CAATAGGCATTTGTTCTACTTAAAAGAAAATCGGAAGCTAGC	BAC A17 BAC A17 BAC A17 BAC A17 BAC A17

NheI

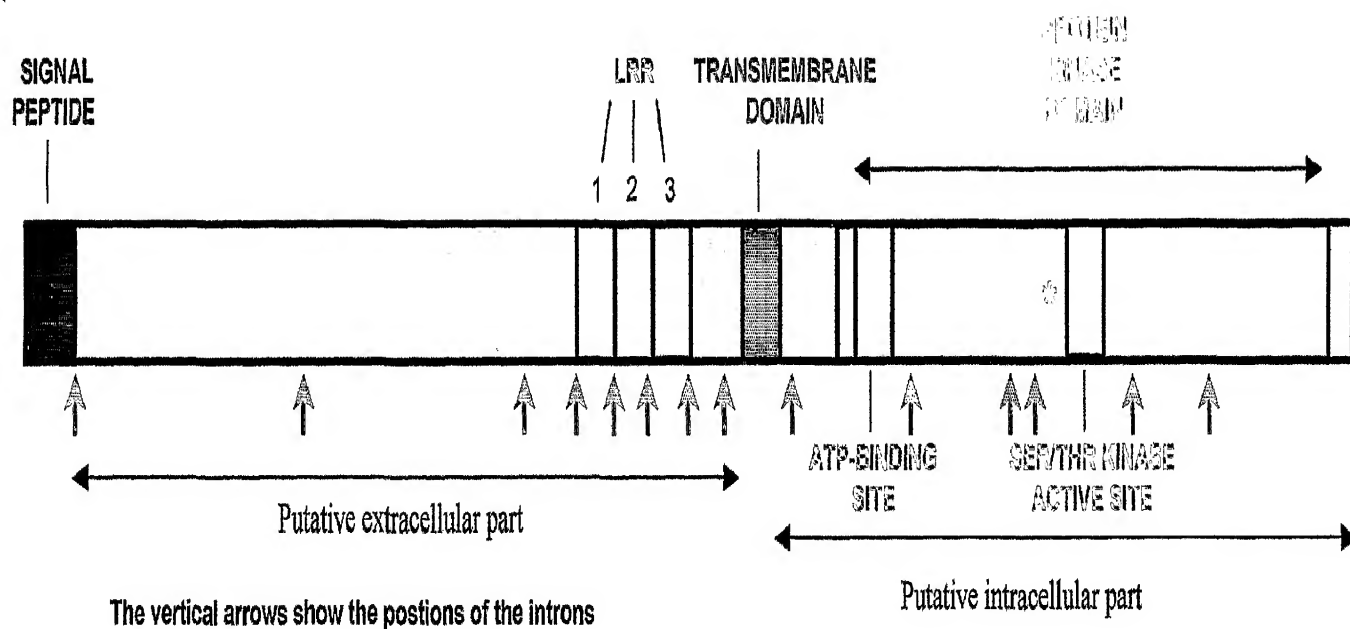
19/60

Figure 12.

MMEHQVIRIFRLVVAFFVLCLCIFIRSASSATKGFESIACCADSNYTDEPKTTLTYYTTDHIW
FSDKRSCRQIPEILFSHRSNKNVRKFEIYEGKRCYNLPTVKDQVYLIRGIFPFDSLNSSF
YVSIGVTELGE LRSSRLLEDLEIEGVFRATKDYIDFCLLKEDVNPFTISQIELRPLPEEYLH
GFGTSVLKLI SRNNLGD TNDDIRFPDDQNDRIWKRKETSTPTSALPLSFNVSNDLKDSV
TPPLQVLQTALTHPERLEFVHDGLETDDYEYSVFLHFLELNGTVRAGQRFVDIYLNNEIK
KEKFDVLAGGSKNSY TALNISANGSLNITLVKASGSEFGPLL NAYEILQARSWIETNOK
DLEVIQKMREELLLHNQENEALSWSGDPCMIFPWKGITCDDSTGSSIITKLDLSSNNLK
GAIPSIIVTKMTNLQILNLSHNQFDMLFPSFPSSLLISLDLSYNDLSGWLPEIIISLPHL
KS LYFGCNPMSDEDTTKLNSSLINTDYGRCKAKKPKFGQVFVIGAITSGSLITLAVGI
LFFCRYRHKSITLEGFGKTYPMATNII FSLPSKDDFFIKSVSVKPF TLEYIEQATEQYKT
LIGEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGTREFDNELNLLSAIQHENLVPLLGYCN
EYDQQILVYPFMSNGSLLDRLYGEASKRKILDWPTRL SIALGAARGLAYLHTFPGRSVIH
RDVKSSNILLDQSMCAKVADFGFSKYAPQEGDSYVSLEVRGTAGYLDPEYYKTQQLSEKS
DVFSFGVVLLEIVSGREPLNIKRPRIEWSLVEWAKPYIRASKVDEIVDPGIKGGYHAEAL
WRVVEVALQCLEPYSTYRPCMVDIVRELEDALIIENNA SEYMKSIDSLGGSNRY SIVMDK
RALPSTTSTAESTITTQTL SHPQPR*

Figure 13.

A.



LRR = Leucine Rich Repeats

The star (*) indicates the site of the "stop codon" mutation in the NORK gene of the MN1008 plant

20/60

B.

The name of the domain

Position within the protein

Signal peptide

1-28 aa

1. Leucine Rich Repeat (LRR)

405-429 aa

2. Leucine Rich Repeat (LRR)

430-454 aa

3. Leucine Rich Repeat (LRR)

455-476 aa

Transmembrane domain

523-543 aa

ATP binding site

601-623 aa

Serine-Threonine kinase active site

717-729 aa

Putative N-glycosylation sites

43-46; 116-119; 257-260; 318-320; 322-325; 326-329;
436-439; 499-502; 628-631; 673-676; 876-879 aa

Putative phosphorylation (autophosphorylation) sites

61-63; 65-67; 75-77; 78-80; 98-100; 133-135; 168-170;
173-175; 193-195; 222-224; 495-497; 581-583;
685-687; 776-778; 793-795; 855-857; 890-892 aa

21/60
Figure 14.

TTCCAGGCCCTAAAGTCAAACACCATATTTTAAACAATATTCTTTCTTCTACAAGGGTATAA
CTTTTATACAAGTTCACATATATTATAGGATTGATCAAGGTTCAATTTTTCTTTCTTTGAA
AAATCTCTAAGGGGTGCTGTTTCCAAGGCAGAAAATGAAATAGAATTGAGAAGAATTTTT
ATGGTACTAAAGGGAAGATGAGAAGTTAGTTAGCATGGATTCAAATTTGATAACCCCTTG
GGGTAAAATCTCTTTTCAGATTATAATGGAGCTACAAGTTATTAGGATATTTAGATTGGTT
GTGGCATGTGTTCTTTGTTTGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAGGG
TTTGAGAGCATAGCATGTTGTGCTGATTCCAATTACACAGATCCAAAACCACCTAACT
TATACAACAGATCACATCTGGTTCTCTGATAAAAGAAGTTGCAGACCAATACCCGAAAT
TTGTTTAGCCACAGAAGCAATAAAAATGTTTGAATATTTGAAATAGATGAAGGAAAGAGA
TGTATACCTTTGCCAACAATTAAGGATCAAGTATATTTGATAAGGGGTGTATTTCCCTTT
GATAGTTTAAATCTTCGTTTTATGTTTATATCGGGGTAAACAGAACTAGGTGAATTAAGA
TCGTCTAGACTCGAGGACTTGGAAATTGAGGGAGTTTTTAGAGCCACCAGCAATATATT
GATTTCTGCTTATTGAAGGAGGATGTCAATCCCTTCATTTCTCAGATTGAATTGAGGCCA
TTACCTGAAGAATACCTACATGGTTTCGCTACTAGTGTTTTTAAACTGATAAGCAGAAAC
AATCTTGGTGACATAAATGATGATATCAGGTTCCCAGATGACCGAAATGATAGAATCTGG
AAACGGAAAGCAACTTCAACTCCATCATCTGCCCTTCCACTGTCTTTCAATGTCAGCAAT
GTTGACCTCAAAGACAGTGTGCGACCTCCTCTACAAGTCTTACAACAGCTCTTACTCAC
CCTGAGCGATTGGAGTTTGTCCATGATGGCTCGAGACCGATGATTATGAATACTCTGTG
TTTCTCCACTTTCTTGAATAAATGGCACTGTCAGAGCAGGACAAAGGGTGTGTGACATC
TATCTAAACAATGAGATTAAAAAGGAGAAGTTTGATGTTTTGGCTGGAGGGTCCAAGAAC
AGTTACACTGCCCTTGACATTTTCAAGCAATGGATCACTCAATATAACCTTAGTCAAGGCA
TCTGGATCTGAGTTTGGACCCCTTTTGAATGCCATGAAATCCTGCAGGCACGGTCGTGG
ATTGAAGAGACCAACCAAAAAGATTTGGAACCTATTTCAGAAGACAAGAGAAGAAGTCTG
CTGCACAACCAAGAAAATGAAGCATTGGAGAGTTGGAGTGGAGACCCCTTGATGATTTTC
CCCTGGAAAGGAATAACATGTGATGATTCAACTGGTTCATCTATTATCACTATGCTGGAT
CTTTCTTCCAATAATCTCAAGGGAGCAATTCCTTACTTTGTCACTAAGATGACCAATTTA
CAAATACTGAACCTGAGCCACAACCAGTTCGATTGTTTATCCCTCGTTTCCACCGTCC
TCCTTGCTGATATCATTGGATCTGAGCTACAATGATCTTGATGGACGGCTTCCAGAATCC
ATTATCTCACTGCCACATTTAAATCATTATATTTTGGCTGCAATCCATATATGAAGGAC
GAAGATACAACAAAGTTGAACAGTTCACTAATCAATACAGATTATGGGAGATGCAAAAGGA
AAAAAACCAGTTTGGACAAGTATTCGTGATTGGAGCTATTACAAGTGATCACTTTTG
ATTACTTTGGCTGTTGGAATTTCTATTTTTTTGCGGTTATAGACACAAGTCAATTACTTTG
GAAGGATTTGGTGGAAGACCTACCCAATGGCAACAAATATAATCTTCTCTTTGCCAAGC
AAAGACGATTTCTTCATAAAGTCTGTATCAGTTAAACCATTCACTTTGGAGTATATAGAG
CAGGCTACAGAACAGTACAAAACCTTTGATATGTGAAGGAGGATTTGGTTCTGTTTACAGA
GGCACTCTAGACGATGGTCAAGAAGTGGCAGTGAAAGTGCGGTTCATCCACATCAACTCAG
GGAACCAAGGAATTTGATAACGAGCTAAACCTACTTTTCACTATACAACATGAGAACCTG
GTGCCTCTTCTGGGTTACTGTAATGAGTATGATCAACAAATTTCTCGTGTATCCATTCTATG
TCTAATGGCTCTTTGCTAGATAGATTATACGGGGAAGCATCAAAGAGAAAAATATTAGAC
TGGCCAACCTAGACTCTCTATTGCTCTCGGTGCACCTCGAGTTTGGCATAGCTTTCACACA
TTTCCAGGACCTTCTGTAATACACAGGGACCTAAAATCGAGCAATATACTGCTGGATCAG
AGCATGTGTGCTAAGGTTGCAGATTTTGGTTTCTCAAAAATACGCTCCTCAGGAAGGAGAC
AGTTATGTTTTCCCTTGAAGTAAGAGGAACTGCAGGGTATCTGGATCCTGAGTACTACAAA
ACCCAGCAATTATCTGAAAAAAGTGATGTTTTTCAAGCTTTGGTGTGGTTCTACTTGAAATT
GTAAGTGGACGGGAACCTCTCAACATAAAGAGACCACGGATCGAGTGGAGCTTGGTTGAA
TGGGCTAAACCATAACATAAGAGCATCAAAGGTGGATGAAATTGTAGATCCTGGCATCAAG
GGAGGATATCATGCAGAAGCATTGTGGAGAGTTGTGGAAGTAGCACTGCAATGTCTAGAA
CCCTACTCAACATATAGGCCATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTC
ATTATTGAAAACAATGCATCTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAAC
CGTACTCAATTGTTATGGACAAACGGGCGCTGCCTTCAACTACATCTACAGCAGATCA
ACTATCACAAACCAACCTTGACACACCTCAACCGAGATAGTAAATGGGTGATGGAAT
TCTTTTGATTTGTTTTTGATCATTGCTTTAGTAAATATCCATTTTCAATGGGTAAAGGAGA
AAAAATACTACTTTTGATTGTATTTTTCATCCACTCTATGTTTTCTTGAACTGAATCTCTCT
TGCTCAGCCCCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAAACCA
TATGGTGCATAATTTGAAAGCCATATTATATCATTTGCCAAGTCCAAAGTAAAAATTTCA
CAAACCTAGTTAGATTGCGATTTAGTCTATAGACACTTCAACAGAGCTATATACACTATGG
TTGACTTGCGACTAATTC

22/60
Figure 15.

CCATATTTTAAACAATATTCTTTCTTCTACAAGGGTATAACTTTTATACAAGTTCACTATA
TTATAGGATTGATCAAGGTCATTTTTTCTTTCTTTGAAAAATCTCTAAGGGGTGTGGTT
TCCAAGGCAGAAAATGAAATAGAATGCAGAAGAATTTGTATGGTACTATAAAGGGAAGAT
GAAAAGTTAGTTAGCATGGATTCAAGTTTGATAACCCTTTGGGGTAAAAATCTCTTTTCAGA
TTATGATGGAGCTACAAGTTATTAGGATATTTAGATTGGTTGTGGCATGTGTTCTTTGTT
TGTGTATATTTATCAGATCAGCTTCTTCTGCAACTAAAGGGTTTGAGAGCATATCATGTT
GTGCTGATTCCAATTACACAGATCCAAAAACAACCCTAACTTATACAACAGATCACATCT
GGTTCCTCTGATAAAAGAAGTTGCAGACCAATACCCGAAATTTTGTGTTAGCCACAGAAGCA
ATAAAAATGTTTGAATATTTGAAATAGATGAAGGAAAGAGATGTTATACTTTGCCAACAA
TTAAGGATCAAGTATATTTGATAAGGGGTGTATTTCCCTTTGATAGTTTAAATCTTCGT
TTTATGTTTATATCGGGGTAAACAGAACTAGGTGAATTAAGATCGTCTAGACTCGAGGACT
TGGAAATCGAGGGAGTTTTTAGAGCCACCAAAGACTATATTGATTTCTGCTTATTGAAGG
AAGATGTCAATCCCTTCATTTCTCAGATTGAATTGAGGCCATTACCTGAAGAATACCTAC
ATGGTTTCGCTACTAGTGTTTTAAAACTGATAAGCAGAAATAATCTTGTTGACACAAATG
ATGATATAAGGTTCCAGATGACCAAAATGATAGAATCTGGAAACGGAAAGCAACTTCAA
CTCCATCATCTGCCCTTCCCCTGTCTTCCAATGTCAGCAATGTTGACCTCAAAGACAGTG
TCACACCTCCTCTACAAGTCCTACAAACAGCTCTTACTCACCTGAGCGATTGGAGTTCG
TCCATGATGGCCTCGAGACCGATGATTATGAATACTCTGTGTTTCTCCACTTCTTGAAC
TAAATGGCACTGTGAGAGCAGGACAAAGGGTGTGTTGACATCTATCTAAACAATGAGATTA
AAAAGGAGAAGTTTGATGTTTTGGCTGGAGGGTCCAAGAACAGTTACACTGCCTTGAACA
TTTCAGCAAATGGATCACTCAATATAACCTTAGTCAAGGCATCTGGATCTGAGTTTGGAC
CCCTTTTGAATGCCTATGAAATCCTGCAGGCACGGTCTGGATTGAAGAGACCAACCAAA
AAGATTTGGAACCTATTAGAGATGAGAGAAGAAGTCTGCTGCACAACCGAGAAAATG
AAGCATTGGAGAGTTGGAGTGGAGACCTTGTATGATTTTTCCCCTGGAAAGGAATAACAT
GTGATGATTCAACTGGTTCATCTATTATCACTATGCTGGATCTTTCTTCCAATAATCTCA
AGGGAGCAATTCCTTACTTTGTCACTAAGATGACCAATTTACAAATACTGAACCTGAGCC
ACAACCAGTTCGATTCTGTTATTTCCCCTCGTTTCCACCGTCTCCTTGCTGATATCATTTGG
ATCTGAGCTACAATGATCTTGATGGACGGCTTCCAGAATCCATTATCTCACTGCCACATT
TAAAATCATTATATTTTTGGCTGCAATCCATATATGAAGGACGAAGATACAACAAAGTTGA
ACAGTTCACTAATCAATACAGATTATGGGAGATGCAAAGGAAAAAAACCAGTTTGGAC
AAGTATTCGTGATTGGAGCTATTACAAGGGGATCACTTTTGTATTACTTTGGCTGTTGGAA
TTCTATTTTTTTTGGCGTTATAGACACAAGTCAATTACTTTTGGAAAGGATTGTTGGAAAGA
CCTACCCAATGGCAACAAATATAATCTTCTTTGCCAAGCAAAGACGATTTCTTCTATAA
AGTCTGTATCAGTTAAACCATTCACTTTGGAGTATATAGAGCAGGCTACAGAACAGTACA
AACTTTGATAGGTGAAGGAGGATTTGGTTCTGTTTACAGAGGCACTCTAGACGATGGTC
AAGAAGTGGCAGTGAAAGTGCGGTCATCCACATCAACTCAGGGAACCCNAGAAATTTGATA
ATGAGCTAAACCTACTTTGAGCTATAACAATGAGAACCTGGTGCCTCTTCTGGGTTACT
GTAATGAGTATGATCAACAAATCTCGTGTATCCATTGATGTTCAATGGCTCTTTGCTAG
ATAGACTATACGGGGAAGCATCAAAGAGAAAAATATTAGACTGGCCAACCTAGACTCTCTA
TTGCTCTCGGTGCAGCTCGAGGTTTGGCATATCTTACACATTTCCAGGACGTTCTGTAA
TACACAGGGACGTAAAATCGAGCAATATACTGCTGGATCAGAGCATGTGTGCTAAGGTTG
CAGATTTTTGGTTTCTCAAAATACGCTCCTCAGGAAGGAGACAGTTATGTTTCCCTTGAAG
TAAGAGGAACTGCAGGGTATCTGGATCCTGAGTACTACAAAACCCAGCAATTAATCTGAAA
AAAGTGATGTTTTTCACTTTTGGTGTGGTTCTACTTGAAATTGTAAGTGGACGGGAACCTC
TCAACATAAAGAGACCACGGATCGAGTGGAGCTTGGTTGAATGGGCTAAACCATACATAA
GAGCATCAAAGGTGGATGAAATTGTAGATCCTGGCATCAAGGGAGGATATCATGCAGAAG
CATTTGTGGAGAGTTGTGGAAGTAGCACTGCAATGTCTAGAACCCTACTCAACATATAGGC
CATGCATGGTTGATATTGTCCGCGAGTTGGAGGATGCTCTCATTATTGAAAACAATGCAT
CTGAATACATGAAATCCATAGACAGCCTTGGAGGATCCAACCGCTACTCAATTGTTATGG
ACAAACGGGCGCTGCCTTCAACTACATCTACAGCAGAATCAACTATCACAACCCAAACCT
TGACACACCCTCAACCGAGATAGTAAATGGGTGATGGAATTCTTTTGTATTGTTTTTTA
TCATTTGCTTTTAGTAATATCCATTTTAAATGGTAAGGAGAAAAATACTACTTTTGATTG
TATTTTTCATCCACTCTATGTTTCTTGAAACTGAATCTCTCTTGCTCAGCCCCAGTTTTTA
TGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAACCATATGGTGCATAATTTGAAA
GCCATATTATATCATTTGCTAAGTCCAAAGTAAAAATTTACAAACTAGTTAGATTGCGA
TTTAGTCTATACACACTTCAACAGAGCTATATACACTAT

23/60

Figure 16.

NLFQIIMELQVIRIFRLVVACVLCCLCIFIRSASSATKGFESIACCADSNYTDPKTTLTYYT
TDHIWFSDKRSCRPIPEILFSHRSNKNVRIFEIDEGKRCYTLPTIKDQVYLIRGVFPFDS
LNSSFYVYIGVTELGELRSSRLEDLEIEGVFRATKDYIDFCLLKEDVNPFIHQIELRPLP
EEYLLHGFATSVLKLISRNNLGDINDDIRFPDDRNDRIWKRKATSTPSSALPLSFNVSNVD
LKDSVAPPLQVLQTALTHPERLEFVHDGLETDDYEYSVFLHFLELNGTVRAGQRFVDIYL
NNEIKKEKFDVLAGGSKNSYALNISANGSLNITLVKASGSEFGPLLNAYEILQARSWIE
ETNQKDLELIQKTREELLLHNQENEALSWSGDPCMIFFPWKGITCDDSTGSSIITMLDLS
SNNLKGAIPYFVTKMTNLQILNLSHNQFDSLFPSPFPSSLLISLDLSYNDLDGRLPESII
SLPHLKSLYFGCNPYMKDEDTTKLNSSLINTDYGRCKGKKPKFGQVVFVIGAITSGSLLIT
LAVGILFFCRYRHKSITLEGFGGKTYPMATNIIFFSLPSKDDFFIKSVSVKPFTLEYIEQA
TEQYKTLICEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGTKEFDNELNLLSAIQHENLVP
LLGYCNEYDQQILVYPFMSNGSLLDRLYGEASKRKILDWPTRLIALGAPRGLA*

24/60

Figure 17.

NLFQIMMELQVIRIFRLVVACVLCICIFIRSASSATKGFESIACCADSNYTDPKTTLTYYT
TDHIWFSDKRSCRPIPEILFSHRSNKNVRIFEIDEGKRCYTLPTIKDQVYLIRGVFPFDS
LNSSFYVYIGVTELGE LRSSRLEDLEIEGVFRATKDYIDFCLLKEDVNPFFISQIELRPLP
EEYLGHFATSVLKLISRNNLGDTNDDIRFPDDQNDRIWKRKATSTPSSALPLSSNVSNVD
LKDSVTPPLQVLQTALHTPERLEFVHDGLETDDYEYSVFLHFLELNGTVRAGQRVFDIYL
NNEIKKEKFDVLAGGSKNSYTALNISANGSLNITLVKASGSEFGPLL NAYEILQARSWIE
ETNQKDLELIQKMREELLLHNRENEALESWSGDPCMIFPWKGITCDDSTGSSIITMLDLS
SNNLKGAIPYFVTKMTNLQILNLSHNQFDSLFPSPFPSSLLISLDLSYNDLDGRLPESII
SLPHLKSLYFGCNPYMKDEDTTKLNSSLINTDYGRCKGKKPKFGQV FVIGATTRGSLIT
LAVGILFFCRYRHKSI TLEGFGGKTYPMATNII FSLPSKDDFFIKSVSVKPF TLEYIEQA
TEQYKTLIGEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGT XEFDNELNLLSAIQHENLVP
LLGYCNEYDQQILVYPFMSNGSLDLRYGEASKRKILDWPTRL SIALGAARGLAYLHTFP
GRSVIHRDVKSSNILLDQSMCAKVADFGFSKYAPQEGDSYVSLEVRGTAGYLDPEYYKTQ
QLSEKSDVFSFGVLL EIVSGREPLNIKRPIEWSLVEWAKPYIRASKVDEIVDPGIKGG
YHAEALWRVVEVALQCLEPYSTYRPCMVDIVRELEDALIIEN NASEYMKSIDSLGGSNRY
SIVMDKRALPSTTSTAESTITTQTLTHPQPR*

25/60
Figure 18. (1/6)

```

.....TTCCAGGCCTAAAGTCAAACACCATATTTTAAACAATATTCTTTCTTCTAC-----AGCTTTTATACAAGTT Mta17
      TTCCAGGCCTAAAGTCAAACACCATATTTTAAACAATATTCTTTCTTCTACAAGGGTATAACTTTTATACAAGTT Ms1N-
      CCATATTTTAAACAATATTCTTTCTTCTACAAGGGTATAACTTTTATACAAGTT Ms6N+
TTTTTTTAAACAATTTTCTTTTCTAATTTCTAACCCCTCCATCAAATTGAGAACTTTTGTAATAATGGTTTCAAATCCTTT Ps

CACTATATTATAGGATTGATCAGGGTTCATTTTTTCTTTCTTTTGAAAAATCTCTAAGGGGTGCTGTTTCCAAGGCAGAG Mta17
CACTATATTATAGGATTGATCAAGGTTCAATTTTTTCTTTCTTTTGAAAAATCTCTAAGGGGTGCTGTTTCCAAGGCAGAA Ms1N-
CACTATATTATAGGATTGATCAAGGTTCAATTTTTTCTTTCTTTTGAAAAATCTCTAAGGGGTGCTGTTTCCAAGGCAGAA Ms6N+
      CACAAAGGTATAACTTTTGTACATGTTTACTATATTAGAGGATTGATCAAGTTCCCTTTGAAAAATCTCTAGGGGG Ps

AATGAAATAGAATTCAGAAGAATTTTTATGTTACTATAAAGGAAAGATGAAAAGTTAGTTAGCATGGATTCAAGTTTGA Mta17
AATGAAATAGAATTCAGAAGAATTTTTATGGTACT--AAAGGGAAGATGAGAAGTTAGTTAGCATGGATTCAAATTTGA Ms1N-
AATGAAATAGAATTCAGAAGAATTTGTATGGTACTATAAAGGGAAGATGAAAAGTTAGTTAGCATGGATTCAAGTTTGA Ms6N+
      TGAAATAGAATTCAGAAGAATTTTTATGGTACTATA--GGGAAGATGGAGAGTTAGTTAGCATGGATTCCGAGTTTGA Ps

TAACCCTATGGGGTAAAAATCTCTTTTCAGATT ATG ATG GAG TTA CAA GTT ATT AGG ATA TTT AGA TTG Mta17
TAACCCTTTGGGGTAAAAATCTCTTTTCAGATT ATA ATG GAG CTA CAA GTT ATT AGG ATA TTT AGA TTG Ms1N-
TAACCCTTTGGGGTAAAAATCTCTTTTCAGATT ATG ATG GAG CTA CAA GTT ATT AGG ATA TTT AGA TTG Ms6N+
GAACCCTTTGGGGTAAACCTATCTTTTCGGATT ATG ATG GAG CTA CGA GTT ATT TGT ATA ATA AGA TTG Ps

GTT GTG GCA TTT GTT CTT TGT TTG TGT ATA TTT ATC AGA TCA GCT TCT TCT GCA ACT AAA Mta17
GTT GTG GCA TGT GTT CTT TGT TTG TGT ATA TTT ATC AGA TCA GCT TCT TCT GCA ACT AAA Ms1N-
GTT GTG GCA TGT GTT CTT TGT TTG TGT ATA TTT ATC AGA TCA GCT TCT TCT GCA ACT AAA Ms6N+
GTT GTG GCG TGT GTT CTT TGT CTG TGT ATA TTT ATC AGA TCA GCT TCT TCT GCA ACT GAA Ps

G i1 GG TTT GAG AGC ATA GCA TGT TGT GCT GAT TCA AAT TAC ACA GAT CCA AAA ACC ACC Mta17
G i1 GG TTT GAG AGC ATA GCA TGT TGT GCT GAT TCC AAT TAC ACA GAT CCA AAA ACC ACC Ms1N-
G i1 GG TTT GAG AGC ATA GCA TGT TGT GCT GAT TCC AAT TAC ACA GAT CCA AAA ACA ACC Ms6N+
G i1 GG TTT GAG AGC ATA GCA TGC TGT GCT GAT TCG AA Ps

CTA ACT TAT ACA ACA GAT CAC ATC TGG TTC TCT GAT AAA AGA AGT TGC AGA CAA ATA CCC Mta17
CTA ACT TAT ACA ACA GAT CAC ATC TGG TTC TCT GAT AAA AGA AGT TGC AGA CCA ATA CCC Ms1N-
CTA ACT TAT ACA ACA GAT CAC ATC TGG TTC TCT GAT AAA AGA AGT TGC AGA CCA ATA CCC Ms6N+

GAA ATT TTG TTT AGC CAC AGA AGC AAT AAA AAT GTT CGA AAA TTT GAA ATA TAT GAA GGA Mta17
GAA ATT TTG TTT AGC CAC AGA AGC AAT AAA AAT GTT CGA ATA TTT GAA ATA GAT GAA GGA Ms1N-
GAA ATT TTG TTT AGC CAC AGA AGC AAT AAA AAT GTT CGA ATA TTT GAA ATA GAT GAA GGA Ms6N+

AAG AGA TGT TAT AAT TTG CCA ACA GTT AAG GAT CAA GTA TAT TTG ATA AGG GGC ATA TTT Mta17
AAG AGA TGT TAT ACT TTG CCA ACA ATT AAG GAT CAA GTA TAT TTG ATA AGG GGT GTA TTT Ms1N-
AAG AGA TGT TAT ACT TTG CCA ACA ATT AAG GAT CAA GTA TAT TTG ATA AGG GGT GTA TTT Ms6N+

CCC TTT GAT AGT TTA AAT TCT TCG TTT TAT GTT TCG ATC GGG GTA ACA GAA CTA GGC GAA Mta17
CCC TTT GAT AGT TTA AAT TCT TCG TTT TAT GTT TAT ATC GGG GTA ACA GAA CTA GGT GAA Ms1N-
CCC TTT GAT AGT TTA AAT TCT TCG TTT TAT GTT TAT ATC GGG GTA ACA GAA CTA GGT GAA Ms6N+

TTA AGA TCG TCT AGG CTC GAG GAC TTG GAA ATT GAG GGA GTT TTT AGA GCC ACC AAA GAC Mta17
TTA AGA TCG TCT AGA CTC GAG GAC TTG GAA ATT GAG GGA GTT TTT AGA GCC ACC AAA GAC Ms1N-
TTA AGA TCG TCT AGA CTC GAG GAC TTG GAA ATC GAG GGA GTT TTT AGA GCC ACC AAA GAC Ms6N+
T Vv

TAC ATA GAT TTC TGC TTA TTG AAG GAG GAT GTC AAT CCC TTC ATT TCT CAG ATT GAA TTG Mta17
TAT ATT GAT TTC TGC TTA TTG AAG GAG GAT GTC AAT CCC TTC ATT TCT CAG ATT GAA TTG Ms1N-
TAT ATT GAT TTC TGC TTA TTG AAG GAA GAT GTC AAT CCC TTC ATT TCT CAG ATT GAA TTG Ms6N+
GAC ATA GAC TTC TGC TTA TTG AAG GAG GAT GCC AAT CCC TTC ATT TCT CAG CTT GAA CTG Vv

```

26/60

Figure 18. (cont. 2/6)

AGG CCA TTA CCT GAA GAA TAC CTA CAT GGT TTC GGT ACT AGT GTT TTA AAA CTG ATA AGC	MtA17
AGG CCA TTA CCT GAA GAA TAC CTA CAT GGT TTC GCT ACT AGT GTT TTA AAA CTG ATA AGC	Ms1N-
AGG CCA TTA CCT GAA GAA TAC CTA CAT GGT TTC GCT ACT AGT GTT TTA AAA CTG ATA AGC	Ms6N+
AGG CCA TTA CCT GAA GAA TAC ATG CAT GAT TTC TCT ACC AGC GTT TTA AAA CTG ATA AAC	Vv
AGA AAC AAT CTT GGT GAC ACA AAT GAT GAT ATA AG i2 G TTC CCA GAT GAC CAA AAT GAT	MtA17
AGA AAC AAT CTT GGT GAC ATA AAT GAT GAT ATC AG i2 G TTC CCA GAT GAC CGA AAT GAT	Ms1N-
AGA AAT AAT CTT GGT GAC ACA AAT GAT GAT ATA AG i2 G TTC CCA GAT GAC CAA AAT GAT	Ms6N+
AGA AAT AAT CTT TGT GGC ATA GAA GAC GAC ATC AG i2 G TTC CCT GTT GAC CAA AAT GAT	Vv
AGA ATC TGG AAA CGG AAA GAA ACT TCA ACT CCA ACA TCT GCC CTT CCA CTG TCT TTC AAT	MtA17
AGA ATC TGG AAA CGG AAA GCA ACT TCA ACT CCA TCA TCT GCC CTT CCA CTG TCT TTC AAT	Ms1N-
AGA ATC TGG AAA CGG AAA GCA ACT TCA ACT CCA TCA TCT GCC CTT CCC CTG TCT TCC AAT	Ms6N+
AGA ATC TGG AAA --- --- GCA ACT TCA ACT CCA TCA TAT GCT GTT CCA CTG TCT TTC AAT	Vv
GTC AGC AAT GTT GAC CTC AAA GAC AGT GTC ACA CCT CCT CTA CAA GTC CTA CAA ACA GCT	MtA17
GTC AGC AAT GTT GAC CTC AAA GAC AGT GTC GCA CCT CCT CTA CAA GTC CTA CAA ACA GCT	Ms1N-
GTC AGC AAT GTT GAC CTC AAA GAC AGT GTC ACA CCT CCT CTA CAA GTC CTA CAA ACA GCT	Ms6N+
GTC AGT GAT GTT GAT CTC AAC GGC AAA GTG ACA CCT CCT CGA CAA GTC CTA CAA ACA GCT	Vv
CTT ACT CAC CCT GAG CGA TTG GAG TTC GTC CAT GAT GGC CTC GAG ACC GAT GAT TAT GAA	MtA17
CTT ACT CAC CCT GAG CGA TTG GAG TTT GTC CAT GAT GGC CTC GAG ACC GAT GAT TAT GAA	Ms1N-
CTT ACT CAC CCT GAG CGA TTG GAG TTC GTC CAT GAT GGC CTC GAG ACC GAT GAT TAT GAA	Ms6N+
CTT ACT CAC CCT GAT CGA TTG GTG TTC GTC CAC GAC GGC CTC GAG ACC GAT GAT TAT GAA	Vv
TAC TCT GTG TTT CTC CAC TTT CTT GAA CTA AAT GGC ACT GTC AGA GCA GGA CAA AGG GTG	MtA17
TAC TCT GTG TTT CTC CAC TTT CTT GAA CTA AAT GGC ACT GTC AGA GCA GGA CAA AGG GTG	Ms1N-
TAC TCT GTG TTT CTC CAC TTT CTT GAA CTA AAT GGC ACT GTC AGA GCA GGA CAA AGG GTG	Ms6N+
TAC TCT GTG TTA CTC TAC TTT CTT GAA CTA AAT AAC ACT GTC AAA GCA GGA CAA AGG GTG	Vv
TTT GAC ATC TAT CTA AAC AAT GAG ATT AAA AAG GAG AAA TTT GAT GTT TTG GCT GGA GGG	MtA17
TTT GAC ATC TAT CTA AAC AAT GAG ATT AAA AAG GAG AAG TTT GAT GTT TTG GCT GGA GGG	Ms1N-
TTT GAC ATC TAT CTA AAC AAT GAG ATT AAA AAG GAG AAG TTT GAT GTT TTG GCT GGA GGG	Ms6N+
TTT GAC ATC TAT CTA AAC AGT GAG ATT AAA AAG GAG AGT TTT GAT GTA TCG GAA GGA GGA	Vv
TTT GAT GTA TTG GAA GGA GGA	Ps
TCC AAG AAC AGT TAC ACT GCC TTG AAC ATT TCA GCA AAT GGA TCA CTC AAT ATA ACC TTA	MtA17
TCC AAG AAC AGT TAC ACT GCC TTG AAC ATT TCA GCA AAT GGA TCA CTC AAT ATA ACC TTA	Ms1N-
TCC AAG AAC AGT TAC ACT GCC TTG AAC ATT TCA GCA AAT GGA TCA CTC AAT ATA ACC TTA	Ms6N+
TCC AAG TAC AGT TAC ATC ACC TTG AAC ATT TCA GCA AAT GGA TCA CTC AAT ATA ACC TTA	Vv
TCC AAG TAC AGT TAC ACT GTC TTG AAC ATT TCA GCA AAT GGA TCA CTC AAT ATA ACC TTA	Ps
GTC AAG GCA TCT GGA TCT GAG TTT GGA CCC CTT TTG AAT GCC TAT GAA ATC CTG CAG GCA	MtA17
GTC AAG GCA TCT GGA TCT GAG TTT GGA CCC CTT TTG AAT GCC TAT GAA ATC CTG CAG GCA	Ms1N-
GTC AAG GCA TCT GGA TCT GAG TTT GGA CCC CTT TTG AAT GCC TAT GAA ATC CTG CAG GCA	Ms6N+
GCC AAG GCA TCT GGA TCT AAG TTT GGA CCC CTT TTG AAT GCT TAT GAA ATC CTG CAG GCG	Vv
GTC AAG GCA TCT GGA TCT AAG TTT GGA CCC CTT TTG AAT GCC TAT GAG ATC CTG CAG GCG	Ps
CGG TCG TGG ATT GAA GAG ACC AAC CAA AAA GAT T i3 TG GAA GTT ATT CAG AAG ATG AGA	MtA17
CGG TCG TGG ATT GAA GAG ACC AAC CAA AAA GAT T i3 TG GAA CTT ATT CAG AAG ACA AGA	Ms1N-
CGG TCG TGG ATT GAA GAG ACC AAC CAA AAA GAT T i3 TG GAA CTT ATT CAG AAG ATG AGA	Ms6N+
CGA CCA TGG ATC GAT GAG ACC GAC CAA ACA GAT G i3 TG GAA GTT ATT CAG AAG TTG AGA	Vv
CGA CCA TGG ATC GAT GAG ACC GAC CAA ACA GAT C i3 TG GAA GTT ATT CAG AAG ATG AGA	Ps

27/60

Figure 18. (cont. 3/6)

GAA	GAA	CTG	CTG	CTG	CAC	AAC	CAA	GAA	AAT	GAA	GCA	TTG	GAG	AGT	TGG	AGT	GGA	GAC	CCT	MtA17	
GAA	GAA	CTG	CTG	CTG	CAC	AAC	CAA	GAA	AAT	GAA	GCA	TTG	GAG	AGT	TGG	AGT	GGA	GAC	CCT	Ms1N-	
GAA	GAA	CTG	CTG	CTG	CAC	AAC	CGA	GAA	AAT	GAA	GCA	TTG	GAG	AGT	TGG	AGT	GGA	GAC	CCT	Ms6N+	
AAA	GAA	CTA	CTG	CTG	CAA	AAC	CAA	GAC	AAT	GAA	GCA	TTG	GAG	AGT	TGG	AGT	GGA	GAT	CCT	Vv	
AAA	GAA	CTG	CTG	CTG	CAA	AAC	CAA	GAC	AAT	GAA	GCA	TTG	GAG	AGT	TGG	AGT	GGA	GAT	CCT	Ps	
TGT	ATG	ATT	TTC	CCC	TGG	AAA	GGA	ATA	ACA	TGT	GAT	GAT	TCA	ACT	GGT	TCA	TCT	ATT	ATC	MtA17	
TGT	ATG	ATT	TTC	CCC	TGG	AAA	GGA	ATA	ACA	TGT	GAT	GAT	TCA	ACT	GGT	TCA	TCT	ATT	ATC	Ms1N-	
TGT	ATG	ATT	TTC	CCC	TGG	AAA	GGA	ATA	ACA	TGT	GAT	GAT	TCA	ACT	GGT	TCA	TCT	ATT	ATC	Ms6N+	
TGT	ATG	CTT	TTT	CCC	TGG	AAA	GGA	GTA	GCA	TGC	GAT	GGT	TCA	AAT	GGT	TCG	TCT	GTC	ATC	Vv	
TGT	ATG	CTT	TTT	CCA	TGG	AAA	GGA	GTA	GCA	TGT	GAT	GGT	TCA	AAT	GGT	TCG	TCT	GTC	ATC	Ps	
ACT	AAG	CT	i4	G	GAT	CTT	TCT	TCC	AAT	AAT	CTC	AAG	GGA	GCA	ATT	CCT	TCC	ATT	GTC	ACT	MtA17
ACT	ATG	CT	i4	G	GAT	CTT	TCT	TCC	AAT	AAT	CTC	AAG	GGA	GCA	ATT	CCT	TAC	TTT	GTC	ACT	Ms1N-
ACT	ATG	CT	i4	G	GAT	CTT	TCT	TCC	AAT	AAT	CTC	AAG	GGA	GCA	ATT	CCT	TAC	TTT	GTC	ACT	Ms6N+
ACT	AAG	CT	i4	G	GAT	CTT	TCC	TCC	AGT	AAT	CTC	AAA	GGA	ACA	ATC	CCT	TCC	AGT	GTC	ACT	Vv
ACT	AAG	CT	i4	G	GAT	CTT	TCC	TCC	AGT	AAT	CTC	AAA	GGA	ACA	ATC	CCT	TCC	AGT	GTC	ACT	Ps
AAG	ATG	ACC	AAT	TTA	CAA	ATA	CT	i5	G	AAC	CTG	AGC	CAC	AAC	CAG	TTC	GAT	ATG	TTA	TTC	MtA17
AAG	ATG	ACC	AAT	TTA	CAA	ATA	CT	i5	G	AAC	CTG	AGC	CAC	AAC	CAG	TTC	GAT	TCG	TTA	TTC	Ms1N-
AAG	ATG	ACC	AAT	TTA	CAA	ATA	CT	i5	G	AAC	CTG	AGC	CAC	AAC	CAG	TTC	GAT	TCG	TTA	TTC	Ms6N+
GAG	ATG	ACC	AAA	TTA	CAA	ATA	CT	i5	G	AAC	CTG	AGC	CAC	AAC	CAC	TTC	GAT	GGC	TAT	ATC	Vv
GAG	ATG	ACC	AAA	TTA	CAA	ATA	CT	i5	G	AAC	CTG	AGC	CAC	AAC	CAT	TTC	GAT	GGT	TAT	ATC	Ps
CCC	TCG	TTT	CCA	CCG	TCC	TCC	TTG	CTG	ATA	TCA	TT	i6	G	GAT	CTT	AGC	TAC	AAT	GAT	CTT	MtA17
CCC	TCG	TTT	CCA	CCG	TCC	TCC	TTG	CTG	ATA	TCA	TT	i6	G	GAT	CTG	AGC	TAC	AAT	GAT	CTT	Ms1N-
CCC	TCG	TTT	CCA	CCG	NCC	TCC	TTG	CTG	ATA	TCA	TT	i6	G	GAT	CTG	AGC	TAC	AAT	GAT	CTT	Ms6N+
CCC	TCG	TTT	CCA	TCG	TCT	TCC	TTG	TTG	ATA	TCA	GT	i6	A	GAT	CTG	AGC	TAC	AAT	GAC	CTA	Vv
CCC	TCA	TTT	CCA	CCG	TCC	TCC	TTG	TTG	ATA	TCA	GT	i6	A	GAT	CTG	AGC	TAC	AAT	GAC	CTA	Ps
TCA	GGA	TGG	CTT	CCA	GAA	TCC	ATT	ATC	TCA	CTG	CCA	CAT	TTA	AAA	TCA	TT	i7	A	TAT	TTT	MtA17
GAT	GGA	CGG	CTT	CCA	GAA	TCC	ATT	ATC	TCA	CTG	CCA	CAT	TTA	AAA	TCA	TT	i7	A	TAT	TTT	Ms1N-
GAT	GGA	CGG	CTT	CCA	GAA	TCC	ATT	ATC	TCA	CTG	CCA	CAT	TTA	AAA	TCA	TT	i7	A	TAT	TTT	Ms6N+
ACG	GGA	CAG	CTT	CCA	GAA	TCC	ATT	ATC	TCA	CTG	CCA	CAT	TTA	AAC	TCA	TT	i7	G	TAT	TTT	Vv
ACG	GGA	CAG	CTT	CCA	GAA	TCC	ATT	ATC	TCA	CTG	CCA	CAT	TTA	AAC	TCA	TT	i7	A	TAT	TTT	Ps
GGC	TGC	AAT	CCA	TCT	ATG	AGT	GAC	GAA	GAT	ACA	ACA	AAG	TTG	AAC	AGT	TCA	CTA	ATC	AAT	MtA17	
GGC	TGC	AAT	CCA	TAT	ATG	AAG	GAC	GAA	GAT	ACA	ACA	AAG	TTG	AAC	AGT	TCA	CTA	ATC	AAT	Ms1N-	
GGC	TGC	AAT	CCA	TAT	ATG	AAG	GAC	GAA	GAT	ACA	ACA	AAG	TTG	AAC	AGT	TCA	CTA	ATC	AAT	Ms6N+	
GGT	TGC	AAT	CAA	CAC	ATG	AGC	GAC	GAT	GAT	GAA	GCC	AAA	TTG	AAC	AGT	TCA	CTA	ATC	AGT	Vv	
GGC	TGC	AAT	CAA	CAC	ATG	AGA	GAC	GAT	GAT	GAA	GCC	AAA	TTG	AAC	AGT	TCA	CTA	ATC	AAT	Ps	
ACA	GA	i8	T	TAT	GGG	AGA	TGC	AAA	GCA	AAA	AAA	CCA	AAG	TTT	GGA	CAA	GTA	TTC	GTG	ATT	MtA17
ACA	GA	i8	T	TAT	GGG	AGA	TGC	AAA	GGA	AAA	AAA	CCA	AAG	TTT	GGA	CAA	GTA	TTC	GTG	ATT	Ms1N-
ACA	GA	i8	T	TAT	GGG	AGA	TGC	AAA	GGA	AAA	AAA	CCA	AAG	TTT	GGA	CAA	GTA	TTC	GTG	ATT	Ms6N+
ACA	GA	i8	T	TAT	GGG	AGA	TGC	AAG	GCA	AAA	AGC	CCC	AAA	TTT	GGA	CAA	GTA	TTC	ATG	ATT	Vv
ACA	GA	i8	T	TAT	GGG	AGA	TGT	AAT	GCA	AAA	AAA	CCC	AAA	TTT	GGA	CAA	GTA	TTC	ATG	ATT	Ps
GGA	GCT	ATT	ACA	AGT	GGA	TCA	CTT	TTG	ATT	ACT	TTG	GCT	GTT	GGA	ATT	CTA	TTT	TTT	TGC	MtA17	
GGA	GCT	ATT	ACA	AGT	GGA	TCA	CTT	TTG	ATT	ACT	TTG	GCT	GTT	GGA	ATT	CTA	TTT	TTT	TGC	Ms1N-	
GGA	GCT	ATT	ACA	AGT	GGA	TCA	CTT	TTG	ATT	ACT	TTG	GCT	GTT	GGA	ATT	CTA	TTT	TTT	TGC	Ms6N+	
GGA	GCA	ATT	ACA	AGT	GGA	TCG	ATT	TTG	ATT	ACT	TTG	GCT	GTT	GGA	ATT	CTA	TTT	TTT	TGC	Vv	
GGA	GCT	ATT	ACA	AGT	GGA	TCA	ATT	TTG	ATT	ACT	TTG	GCT	GTT	GTA	ATT	CTA	TTC	TTT	TGC	Ps	
CGT	TAT	AGA	CAC	AAG	TCA	ATT	ACT	TTG	GAA	GGA	TTT	GG-	--A	AAG	ACC	TAC	CCA	ATG	GCA	MtA17	
CGT	TAT	AGA	CAC	AAG	TCA	ATT	ACT	TTG	GAA	GGA	TTT	GGT	GGA	AAG	ACC	TAC	CCA	ATG	GCA	Ms1N-	
CGT	TAT	AGA	CAC	AAG	TCA	ATT	ACT	TTG	GAA	GGA	TTT	GGT	GGA	AAG	ACC	TAC	CCA	ATG	GCA	Ms6N+	
CGT	TAT	AGA	CAC	AAG	TCA	ATT	ACT	TTG	GAA	GGA	TTT	GGT	GGA	AAG	ACC	TAC	CCG	ATG	GCA	Vv	
CGT	TAT	AGA	CAC	AAG	TCA	ATT	ACT	TTG	GAA	GGA	TTT	GGT	GGA	AAG	ACC	TAC	CCA	ATG	GCA	Ps	

28/60

Figure 18. (cont. 4/6)

ACA A i9 AT ATA ATT TTC TCT TTG CCA AGC AAA GAC GAT TTC TTC ATA AAG TCT GTA TCA	MtA17
ACA A i9 AT ATA ATC TTC TCT TTG CCA AGC AAA GAC GAT TTC TTC ATA AAG TCT GTA TCA	Ms1N-
ACA A i9 AT AFA ATC TTC TCT TTG CCA AGC AAA GAC GAT TTC TTC ATA AAG TCT GTA TCA	Ms6N+
ACA A i9 AT ATA ATT TTC TCC TTG CCA AGC AAA GAC GAT TTC TTC ATA AAG TCT GTA TCA	Vv
ACA A i9 AT ATA ATT TTT TCC TTG CCA AGC AAA GAC GAT TTC TTC ATA AAG TCT GTA TCA	Ps
GTT AAA CCG TTC ACT TTG GAG TAT ATA GAG CAG GCT ACA GAA CAG TAC AAA ACT TTA ATA	MtA17
GTT AAA CCA TTC ACT TTG GAG TAT ATA GAG CAG GCT ACA GAA CAG TAC AAA ACT TTG ATA	Ms1N-
GTT AAA CCA TTC ACT TTG GAG TAT ATA GAG CAG GCT ACA GAA CAG TAC AAA ACT TTG ATA	Ms6N+
GTT AAA CCG TTC TCT TTG GAG TAT ATA GAG TTG GCA ACA GAG AAG TAC AAA ACT TTG ATA	Vv
GTT AAG CCA TTC ACT TTG GAG TAT ATA GAG TTG GCA ACA GAG AAG TAC AAA ACT TTG ATA	Ps
GGT GAA GGA GGA TTT GGC TCT GTT TAC AGG GGC ACT CTA GAC GAT GGT CAA GAA GTG GCA	MtA17
TGT GAA GGA GGA TTT GGT TCT GTT TAC AGA GGC ACT CTA GAC GAT GGT CAA GAA GTG GCA	Ms1N-
GGT GAA GGA GGA TTT GGT TCT GTT TAC AGA GGC ACT CTA GAC GAT GGT CAA GAA GTG GCA	Ms6N+
GGT GAA GGA GGG TTT GGC TCT GTT TAC CGA GGC ACT CTA GAC GAT GGT CAA GAA GTG GCA	Vv
GGT GAA GGA GGA TTT GGC TCT GTT TAC CGG GGC ACT CTA GAC GAT GGA CAA GAA GTG GCA	Ps
GTG AAA GTG CGG TCA TCC ACA TCA ACT CAG GGA ACC CGA GAA TTT GAT AAT GAG i10 CTA	MtA17
GTG AAA GTG CGG TCA TCC ACA TCA ACT CAG GGA ACC AAG GAA TTT GAT AAC GAG i10 CTA	Ms1N-
GTG AAA GTG CGG TCA TCC ACA TCA ACT CAG GGA ACC CGA GAA TTT GAT AAT GAG i10 CTA	Ms6N+
GTG AAA GTC CGG TCA GCC ACA TCA ACT CAG GGA ACC CGA GAA TTT GAC AAT GAG i10 CTA	Vv
GTG AAA GTC CGG TCA GCC ACA TCA ACT CAG GGA ACC CGA GAA TTT GAC AAT GAG i10 CTA	Ps
AAC CTA CTT TCA GCT ATA CAA CAT GAG AAC CTG GTG CCT CTT CTG GGT TAC TGT AAT GAG	MtA17
AAC CTA CTT TCA GCT ATA CAA CAT GAG AAC CTG GTG CCT CTT CTG GGT TAC TGT AAT GAG	Ms1N-
AAC CTA CTT TCA GCT ATA CAA CAT GAG AAC CTG GTG CCT CTT CTG GGT TAC TGT AAT GAG	Ms6N+
AAC CTA CTT TCA GCT ATA CAA CAT GAG AAC CTG GTG CCT CTT CTT GGT TAC TGT AAT GAA	Vv
AAC CTA CTT TCA GCT ATA CAA CAT GAG AAC CTG GTG CCT CTT CTT GGT TAC TGT AAT GAA	Ps
TAT GAT CAA CAA ATT CTC GTG TAT CCT TTC ATG TCC AAT GGC TCT TTG CTA GAT AGA CTA	MtA17
TAT GAT CAA CAA ATT CTC GTG TAT CCA TTC ATG TCT AAT GGC TCT TTG CTA GAT AGA TTA	Ms1N-
TAT GAT CAA CAA ATT CTC GTG TAT CCA TTC ATG TCC AAT GGC TCT TTG CTA GAT AGA CTA	Ms6N+
TAT GAT CAA CAA ATT CTT GTG TAC CCT TTC ATG TCT AAT GGA TCT TTG CTA GAT AGA CTA	Vv
TAT GAT CAA CAA ATT CTT GTA TAT CCT TTC ATG TCT AAT GGC TCT TTG CTA TAC AGA CTA	Ps
TAC G i11 GG GAA GCA TCA AAG AGA AAA ATA TTA GAC TGG CCA ACT AGA CTC TCT ATT GCT	MtA17
TAC G i11 GG GAA GCA TCA AAG AGA AAA ATA TTA GAC TGG CCA ACT AGA CTC TCT ATT GCT	Ms1N-
TAC G i11 GG GAA GCA TCA AAG AGA AAA ATA TTA GAC TGG CCA ACT AGA CTC TCT ATT GCT	Ms6N+
TAT G i11 GG GAA CCA GCA AAG AGA AAA ATA CTA GAC TGG CCA ACT AGA CTC TCT ATT GCT	Vv
TAT G i11 GG GAA CCG GCA AAG AGA AAA ATA CTA GAC TGG CCA ACT AGA CTC TCA ATT GCT	Ps
CTC GGT GCA GCT CGA G i12 GT TTG GCA TAT CTT CAC ACA TTT CCA GGA CGT TCT GTA ATA	MtA17
CTC GGT GCA CCT CGA G i12 GT TTG GCA TAG CTT CAC ACA TTT CCA GGA CCT TCT GTA ATA	Ms1N-
CTC GGT GCA GCT CGA G i12 GT TTG GCA TAT CTT CAC ACA TTT CCA GGA CGT TCT GTA ATA	Ms6N+
CTA GGA GCA GCT CGA G i12 GT TTG GCA TAT CTT CAC ACA TTT CCT GGA CGG TCT GTA ATA	Vv
CTA GGA GCA GCT CGA G i12 GT TTG GCA TAT CTT CAC ACG TTT CCT GGA CGA TCT GTA ATA	Ps
mutation resulting in in frame STOP codon	
CAC AGG GAC GTA AAA TCG AGC AAT ATA CTG CTG GAT CAG AGC ATG TGT GCT AAG GTT GCA	MtA17
CAC AGG GAC CTA AAA TCG AGC AAT ATA CTG CTG GAT CAG AGC ATG TGT GCT AAG GTT GCA	Ms1N-
CAC AGG GAC GTA AAA TCG AGC AAT ATA CTG CTG GAT CAG AGC ATG TGT GCT AAG GTT GCA	Ms6N+
CAC AGG GAC GTA AAA TCG AGC AAT ATA CTA CTG GAC CAT AGC ATG TGT GCT AAG GTC GCA	Vv
CAC AGG GAC GTA AAA TCG AGC AAT ATA CTA CTG GAC CAT AGC ATG TGT GCT AAG GTT GCA	Ps

29/60

Figure 18. (cont. 5/6)

GAT	TTT	GGT	TTC	TCA	AAA	TAT	GCT	CCT	CAG	GAA	GGA	GAC	AGT	TAT	GTT	TCC	CTT	GAA	GTA	MtA17
GAT	TTT	GGT	TTC	TCA	AAA	TAC	GCT	CCT	CAG	GAA	GGA	GAC	AGT	TAT	GTT	TCC	CTT	GAA	GTA	Ms1N-
GAT	TTT	GGT	TTC	TCA	AAA	TAC	GCT	CCT	CAG	GAA	GGA	GAC	AGT	TAT	GTT	TCC	CTT	GAA	GTA	Ms6N+
GAT	TTT	GGT	TTC	TCA	AAA	TAC	GCT	CCT	CAG	GAA	GGA	GAC	AGT	TAT	GTT	TCC	CTT	GAA	GTA	Vv
GAT	TTT	GGT	TTC	TCA	AAA	TAC	GCT	CCT	CAG	GAA	GGA	GAC	AGT	TAT	GTT	TCC	CTT	GAA	GTA	Ps
AGA	GGA	ACT	GCA	GGG	TAT	CTG	GAT	CCT	GAG	TAC	TAC	AAA	ACC	CAG	CAA	TTA	TCT	GAA	AAA	MtA17
AGA	GGA	ACT	GCA	GGG	TAT	CTG	GAT	CCT	GAG	TAC	TAC	AAA	ACC	CAG	CAA	TTA	TCT	GAA	AAA	Ms1N-
AGA	GGA	ACT	GCA	GGG	TAT	CTG	GAT	CCT	GAG	TAC	TAC	AAA	ACC	CAG	CAA	TTA	TCT	GAA	AAA	Ms6N+
AGA	GGA	ACT	GCA	GGG	TAT	CTG	GAT	CCC	GAG	TAC	TAC	AAA	ACC	CAG	CAA	TTA	TCT	GAA	AAG	Vv
AGA	GGA	ACC	GCA	GGG	TAT	CTG	GAT	CCT	GAG	TAC	TAC	AAA	ACC	CAG	CAA	TTA	TCT	GAA	AAG	Ps
AGT	GAT	i13	GTT	TTC	AGC	TTT	GGT	GTG	GTT	CTT	CTT	GAA	ATT	GTA	AGC	GGA	CGG	GAA	CCT	MtA17
AGT	GAT	i13	GTT	TTC	AGC	TTT	GGT	GTG	GTT	CTA	CTT	GAA	ATT	GTA	AGT	GGA	CGG	GAA	CCT	Ms1N-
AGT	GAT	i13	GTT	TTC	AGC	TTT	GGT	GTG	GTT	CTA	CTT	GAA	ATT	GTA	AGT	GGA	CGG	GAA	CCT	Ms6N+
AGT	GAT	i13	GTT	TTC	AGC	TTT	GGT	GTG	GTT	CTA	CTA	GAA	ATT	GTA	AGT	GGA	CGG	GAA	CCT	Vv
AGT	GAT	i13	GTT	TTC	AGC	TTT	GGT	GTG	GTT	CTA	CTA	GAA	ATT	GTA	AGT	GGA	CGG	GAA	CCT	Ps
CTC	AAC	ATA	AAG	AGA	CCA	AGG	ATC	GAG	TGG	AGC	TTG	GTT	GAA	TGG	i14	GCT	AAA	CCA	TAC	MtA17
CTC	AAC	ATA	AAG	AGA	CCA	CGG	ATC	GAG	TGG	AGC	TTG	GTT	GAA	TGG	i14	GCT	AAA	CCA	TAC	Ms1N-
CTC	AAC	ATA	AAG	AGA	CCA	CGG	ATC	GAG	TGG	AGC	TTG	GTT	GAA	TGG	i14	GCT	AAA	CCA	TAC	Ms6N+
CTC	AAC	ATA	AAG	AGA	CCA	CGA	GTC	GAG	TGG	AGC	TTG	GTT	GAA	TGG	i14	GCT	AAA	CCA	TAC	Vv
CTC	AAC	ATA	AAG	AGA	CCA	CGA	GTC	GAG	TGG	AGC	TTG	GTT	GAA	TGG	i14	GCT	AAA	CCA	TAT	Ps
ATA	AGA	GCA	TCA	AAG	GTG	GAT	GAA	ATT	GTA	GAT	CCT	GGC	ATC	AAG	GGA	GGA	TAT	CAT	GCA	MtA17
ATA	AGA	GCA	TCA	AAG	GTG	GAT	GAA	ATT	GTA	GAT	CCT	GGC	ATC	AAG	GGA	GGA	TAT	CAT	GCA	Ms1N-
ATA	AGA	GCA	TCA	AAG	GTG	GAT	GAA	ATT	GTA	GAT	CCT	GGC	ATC	AAG	GGA	GGA	TAT	CAT	GCA	Ms6N+
ATA	AGA	GCA	TCA	AAG	GTG	GAT	GAA	ATT	GTA	GAT	CCT	GGC	ATC	AAG	GGA	GGA	TAT	CAT	GCA	Vv
ATA	AGA	GCA	TCA	AAG	GTG	GAT	GAA	ATT	GTA	GAT	CCT	GGC	ATC	AAG	GGA	GGA	TAT	CAT	GCA	Ps
GAG	GCA	TTG	TGG	AGA	GTT	GTG	GAA	GTA	GCA	CTA	CAA	TGT	CTA	GAA	CCC	TAC	TCA	ACA	TAT	MtA17
GAA	GCA	TTG	TGG	AGA	GTT	GTG	GAA	GTA	GCA	CTG	CAA	TGT	CTA	GAA	CCC	TAC	TCA	ACA	TAT	Ms1N-
GAA	GCA	TTG	TGG	AGA	GTT	GTG	GAA	GTA	GCA	CTG	CAA	TGT	CTA	GAA	CCC	TAC	TCA	ACA	TAT	Ms6N+
GAG	GCA	CTG	TGG	AGA	GTG	GTG	GAA	GTA	GCA	CTG	CAA	TGT	CT							Vv
GAG	GCA	CTG	TGG	AGA	GTT															Ps
CGG	CCA	TGC	ATG	GTT	GAT	ATT	GTC	CGC	GAG	TTG	GAG	GAT	GCT	CTC	ATT	ATT	GAA	AAC	AAT	MtA17
AGG	CCA	TGC	ATG	GTT	GAT	ATT	GTC	CGC	GAG	TTG	GAG	GAT	GCT	CTC	ATT	ATT	GAA	AAC	AAT	Ms1N-
AGG	CCA	TGC	ATG	GTT	GAT	ATT	GTC	CGC	GAG	TTG	GAG	GAT	GCT	CTC	ATT	ATT	GAA	AAC	AAT	Ms6N+
GCA	TCT	GAA	TAC	ATG	AAA	TCC	ATA	GAC	AGC	CTT	GGA	GGA	TCC	AAC	CGC	TAC	TCA	ATT	GTT	MtA17
GCA	TCT	GAA	TAC	ATG	AAA	TCC	ATA	GAC	AGC	CTT	GGA	GGA	TCC	AAC	CGC	TAC	TCA	ATT	GTT	Ms1N-
GCA	TCT	GAA	TAC	ATG	AAA	TCC	ATA	GAC	AGC	CTT	GGA	GGA	TCC	AAC	CGC	TAC	TCA	ATT	GTT	Ms6N+
ATG	GAC	AAA	CGG	GCG	CTG	CCT	TCA	ACT	ACA	TCT	ACA	GCA	GAA	TCA	ACT	ATC	ACA	ACC	CAA	MtA17
ATG	GAC	AAA	CGG	GCG	CTG	CCT	TCA	ACT	ACA	TCT	ACA	GCA	GAA	TCA	ACT	ATC	ACA	ACC	CAA	Ms1N-
ATG	GAC	AAA	CGG	GCG	CTG	CCT	TCA	ACT	ACA	TCT	ACA	GCA	GAA	TCA	ACT	ATC	ACA	ACC	CAA	Ms6N+
ACC	TTG	TCA	CAC	CCT	CAA	CCG	AGA	TAG	TAA	ATGGGTCGA--TGGAATTCCTTTTGATTGTTTTTGATCATT									MtA17	
ACC	TTG	TCA	CAC	CCT	CAA	CCG	AGA	TAG	TAA	ATGGGTCGA--TGGAATTCCTTTTGATTGTTTTTGATCATT									Ms1N-	
ACC	TTG	TCA	CAC	CCT	CAA	CCG	AGA	TAG	TAA	ATGGGTCGA--TGGAATTCCTTTTGATTGTTTTTGATCATT									Ms6N+	

30/60

Figure 18. (cont. 6/6)

GCTTTAGTAATATCACATTTTAAATGGTAAAGGAGAAAAATACTACTTCTGATTGTATTTCCATCCACTCTATGTTTCT	MtA17
GCTTTAGTAATATCCCATTTTCAATGGTAAAGGAGAAAAATACTACTTTTGATTGTATTTTCATCCACTCTATGTTTCT	Ms1N-
GCTTTAGTAATATCCCATTTTAAATGGTAAAGGAGAAAAATACTACTTTTGATTGTATTTTCATCCACTCTATGTTTCT	Ms6N+
TGAAACTGAATCTCTCTTGCTCAGCCCCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAACCCATA	MtA17
TGAAACTGAATCTCTCTTGCTCAGCCCCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAACCCATA	Ms1N-
TGAAACTGAATCTCTCTTGCTCAGCCCCAGTTTTTATGGGTGAAAAGAATAATTTGGGTCAAATGCAAGTGAACCCATA	Ms6N+
TGGTGCATAATTTAAAAGCCATATCATATCATTTGCCAAGTCCAAAGTAAAAATTCACAAACTAGTTAGATTGCGATT	MtA17
TGGTGCATAATTTGAAAGCCATATTATATCATTTGCCAAGTCCAAAGTAAAAATTCACAAACTAGTTAGATTGCGATT	Ms1N-
TGGTGCATAATTTGAAAGCCATATTATATCATTTGCTAAGTCCAAAGTAAAAATTCACAAACTAGTTAGATTGCGATT	Ms6N+
TAGTCTATAGACACTTCAACAGAGCTATATACACTATGGTTGACTTGCGA	MtA17
TAGTCTATAGACACTTCAACAGAGCTATATACACTATGGTTGACTTGCGACTAATTC	Ms1N-
TAGTCTATACACACTTCAACAGAGCTATATACACTAT	Ms6N+

31/60

Figure 19. (1/2)

MMELQVIRIFRLVVAFLVLCIFIRSSASSATKGFESIACCADSNYTDPKTTLTYTTDHIW MtA17
 IMELQVIRIFRLVVAFLVLCIFIRSSASSATKGFESIACCADSNYTDPKTTLTYTTDHIW Ms1N-
 MMELQVIRIFRLVVAFLVLCIFIRSSASSATKGFESIACCADSNYTDPKTTLTYTTDHIW Ms6N+
 MELRVICIIIRLVVACVLCIFIRSSASSATEGFESIACCADS Ps

FSDKRSCRPIPEILFSHRSNKNVRKFEIYEGKRCYNLPTVKDQVYLIRGIFPFDSLNSSF MtA17
 FSDKRSCRPIPEILFSHRSNKNVRIFEIDEGKRCYTLPTIKDQVYLIRGVFPFDSLNSSF Ms1N-
 FSDKRSCRPIPEILFSHRSNKNVRIFEIDEGKRCYTLPTIKDQVYLIRGVFPFDSLNSSF Ms6N+

YVSIGVTELGE LRSSRLLEDLEIEGVFRATKDYIDFCLLKEDVNPFISQIELRPLPEEYLH MtA17
 YVYIGVTELGE LRSSRLLEDLEIEGVFRATKDYIDFCLLKEDVNPFISQIELRPLPEEYLH Ms1N-
 YVYIGVTELGE LRSSRLLEDLEIEGVFRATKDYIDFCLLKEDVNPFISQIELRPLPEEYLH Ms6N+
 GIDFCLLKEDANPFISQIELRPLPEEYMH Vv

GFGTSVLKLI SRNNLGD TNDDIRFPDDQNDRIWKRKETSTPTSALPLSFNVS NVDLKDSV MtA17
 GFATSVLKLI SRNNLGD TNDDIRFPDDRNDRIWKRKATSTPSSALPLSFNVS NVDLKDSV Ms1N-
 GFATSVLKLI SRNNLGD TNDDIRFPDDQNDRIWKRKATSTPSSALPLSSNVSNVDLKDSV Ms6N+
 DFSTSVLKLINRNNLCGIEDDIRFPVDQNDRIW--ATSTPSYAVPLSFNVSDVDLNGKV Vv

TPPLQVLQTAL THPERLEFVHDGLETDDYEYSVFLHFLELNGTVRAGQRVFDIYLNNEIK MtA17
 APPLQVLQTAL THPERLEFVHDGLETDDYEYSVFLHFLELNGTVRAGQRVFDIYLNNEIK Ms1N-
 TPPLQVLQTAL THPERLEFVHDGLETDDYEYSVFLHFLELNGTVRAGQRVFDIYLNNEIK Ms6N+
 TPRQVLQTAL THPDRLVFVHDGLETDDYEYSVLLYFLELNNTVKAGQRVFDIYLNSEIK Vv

KEKFDVLAGGSKNSY TALNISANGSLNITLVKASGSEFGPLL NAYEILQARSWI EETNQK MtA17
 KEKFDVLAGGSKNSY TALNISANGSLNITLVKASGSEFGPLL NAYEILQARSWI EETNQK Ms1N-
 KEKFDVLAGGSKNSY TALNISANGSLNITLVKASGSEFGPLL NAYEILQARSWI EETNQK Ms6N+
 KESFDVSEGGSKYSYITL NISANGSLNITLAKASGSKFGPLL NAYEILQARPWIDETDQT Vv
 FDVLEGGSKYSYITV LNISANGSLNITLVKASGSKFGPLL NAYEILQARPWIDETDQT Ps

DLEVIQKMREEL LLHNQENEAL ESWSGDPCMI FFWKGITCDDSTGSSII TKDLSSNNLK MtA17
 DLELIQKTREEL LLHNQENEAL ESWSGDPCMI FFWKGITCDDSTGSSII TMLDLSSNNLK Ms1N-
 DLELIQKMREEL LLHNRENEAL ESWSGDPCMI FFWKGITCDDSTGSSII TMLDLSSNNLK Ms6N+
 DVEVIQKL RKE LLQNQDNEAL ESWSGDPCML FFWKGVACDGSNGSSVITKDLSSNNLK Vv
 DLEVIQKMRKE LLQNQDNEAL ESWSGDPCML FFWKGVACDGSNGSSVITKDLSSNNLK Ps

GAIPSI VTKMTNLQI LNL SHNQFDML FPFSSLLISLDLSYNDLSGWL PESIISLPHL MtA17
 GAIPYFVTKMTNLQI LNL SHNQFDSL FPFSSLLISLDLSYNDLDGRL PESIISLPHL Ms1N-
 GAIPYFVTKMTNLQI LNL SHNQFDSL FPFSSLLISLDLSYNDLDGRL PESIISLPHL Ms6N+
 GTIPSSVTEMTKLQI LNL SHNHFDGYIP SFPPSSLLISVDLSYNDLTGQL PESIISLPHL Vv
 GTIPSSVTEMTKLQI LNL SHNHFDGYIP SFPPSSLLISVDLSYNDLTGQL PESIISLPHL Ps

32/60

Figure 19. (2/2)

KSLYFGCNPSMSDEDTTKLNSSLINTDYGRCKAKKPKFGQVFVIGAITSGSLLITLAVGI MtA17
 KSLYFGCNPYMKDEDTTKLNSSLINTDYGRCKGKKPKFGQVFVIGAITSGSLLITLAVGI Ms1N-
 KSLYFGCNPYMKDEDTTKLNSSLINTDYGRCKGKKPKFGQVFVIGAITSGSLLITLAVGI Ms6N+
 NSLYFGCNQHMSDDDEAKLNSSLISTDYGRCKAKSPKFGQVFMIGAITSGSILITLAVGI Vv
 NSLYFGCNQHMRDDDEAKLNSSLINTDYGRCKNAKKPKFGQVFMIGAITSGSILITLAVVI Ps

LFFCRYRHKSITLEGFG-KTYPMATNII FSLPSKDDFFIKSVSVKPF TLEYIEQATEQYK MtA17
 LFFCRYRHKSITLEGFGGKTYPMATNII FSLPSKDDFFIKSVSVKPF TLEYIEQATEQYK Ms1N-
 LFFCRYRHKSITLEGFGGKTYPMATNII FSLPSKDDFFIKSVSVKPF TLEYIEQATEQYK Ms6N+
 LFFCRYRHKSITLEGFGGKTYPMATNII FSLPSKDDFFIKSVSVKPF SLEYIELATEKYK Vv
 LFFCRYRHKSITLEGFGGKTYPMATNII FSLPSKDDFFIKSVSVKPF TLEYIELATEKYK Ps

TLIGEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGTREFDNELNLLSAIQHENLVPLLGYC MtA17
 TLICEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGTKEFDNELNLLSAIQHENLVPLLGYC Ms1N-
 TLIGEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGTXEFDNELNLLSAIQHENLVPLLGYC Ms6N+
 TLIGEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGTREFDNELNLLSAIQHENLVPLLGYC Vv
 TLIGEGGFGSVYRGTLDDGQEVAVKVRSSSTSTQGTREFDNELNLLSAIQHENLVPLLGYC Ps

NEYDQQILVYPFMSNGSLLDRLYGEASKRKILDWPTRL SIALGAARGLAYLHTFPGRSVI MtA17
 NEYDQQILVYPFMSNGSLLDRLYGEASKRKILDWPTRL SIALGAPRGLA* Ms1N-
 NEYDQQILVYPFMSNGSLLDRLYGEASKRKILDWPTRL SIALGAARGLAYLHTFPGRSVI Ms6N+
 NEYDQQILVYPFMSNGSLLDRLYGEPAKRKILDWPTRL SIALGAARGLAYLHTFPGRSVI Vv
 NEYDQQILVYPFMSNGSLLDRLYGEPAKRKILDWPTRL SIALGAARGLAYLHTFPGRSVI Ps

HRDVKSSNILLDQSMCAKVADFGFSKYAPQEGDSYVSLEVRGTAGYLDPEYYKTQQLSEK MtA17
 HRDVKSSNILLDQSMCAKVADFGFSKYAPQEGDSYVSLEVRGTAGYLDPEYYKTQQLSEK Ms6N+
 HRDVKSSNILLDHSMCAKVADFGFSKYAPQEGDSYVSLEVRGTAGYLDPEYYKTQQLSEK Vv
 HRDVKSSNILLDHSMCAKVADFGFSKYAPQEGDSYVSLEVRGTAGYLDPEYYKTQQLSEK Ps

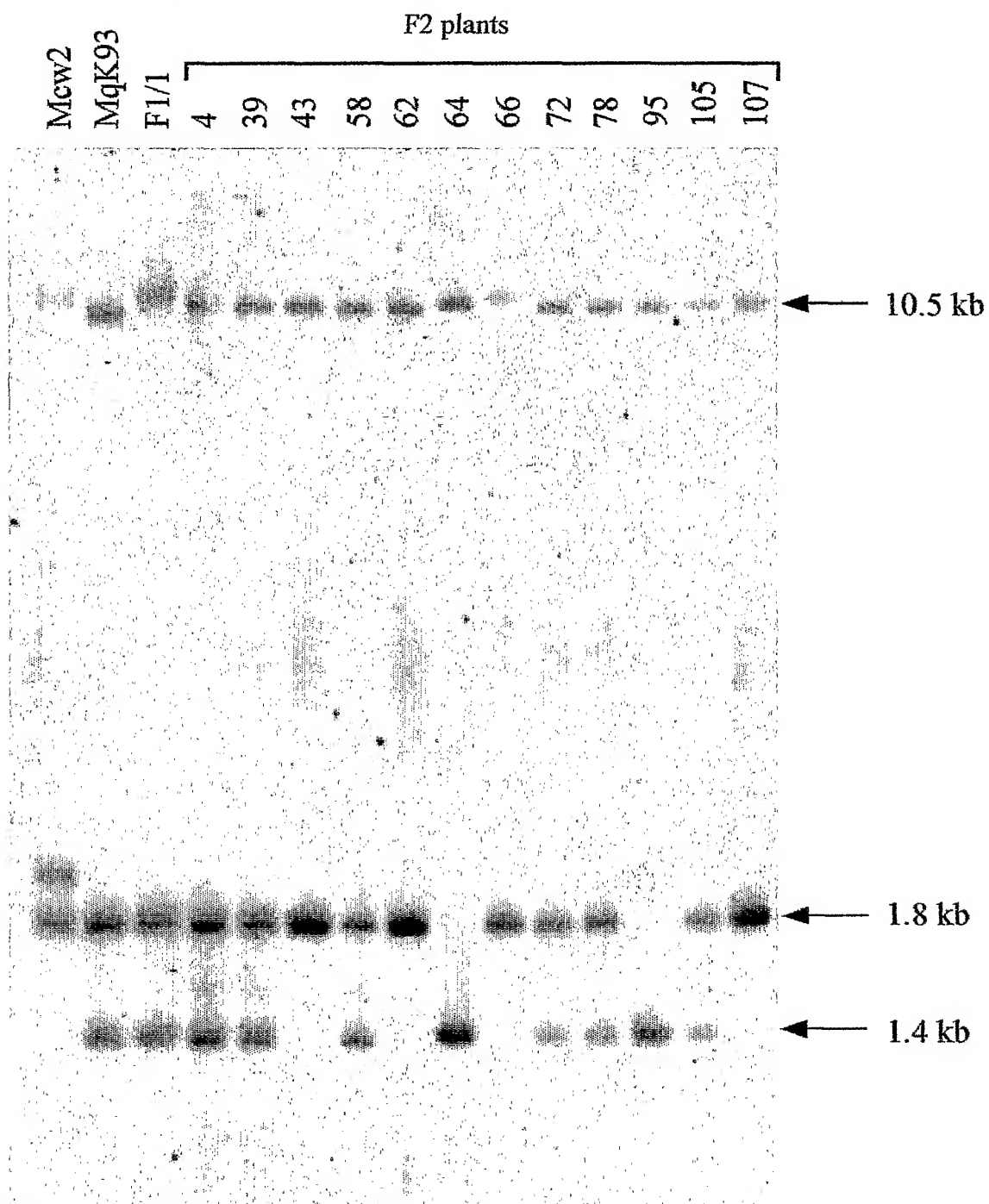
SDVFSFGVVLLEIVSGREPLNIKRPRIEWSLVEWAKPYIRASKVDEIVDPGIKGGYHAEA MtA17
 SDVFSFGVVLLEIVSGREPLNIKRPRIEWSLVEWAKPYIRASKVDEIVDPGIKGGYHAEA Ms6N+
 SDVFSFGVVLLEIVSGREPLNIKRPRIEWSLVEWAKPYIRASKVDEIVDPGIKGGYHAEA Vv
 SDVFSFGVVLLEIVSGREPLNIKRPRIEWSLVEWAKPYIRASKVDEIVDPGIKGGYHAEA Ps

LWRVVEVALQCLEPYSTYRPCMVDIVRELEDALIIENNASEYMKSIDSLGGSNRYISIVMD MtA17
 LWRVVEVALQCLEPYSTYRPCMVDIVRELEDALIIENNASEYMKSIDSLGGSNRYISIVMD Ms6N+
 LWRVVEVALQCL Vv
 LWRV Ps

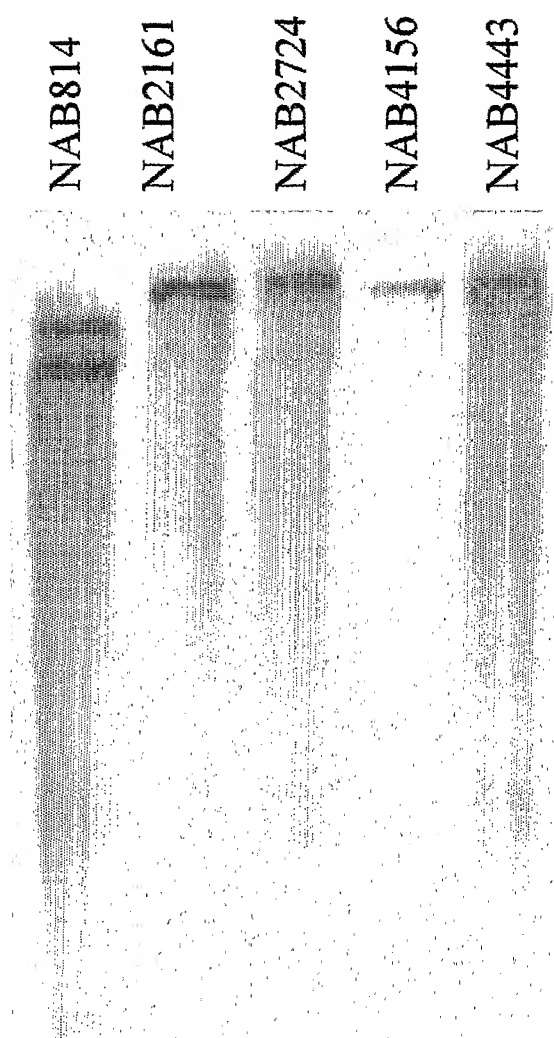
KRALPSTTSTAESTITTQTL SHPQPR** MtA17
 KRALPSTTSTAESTITTQTL THPQPR** Ms6N+

33/60

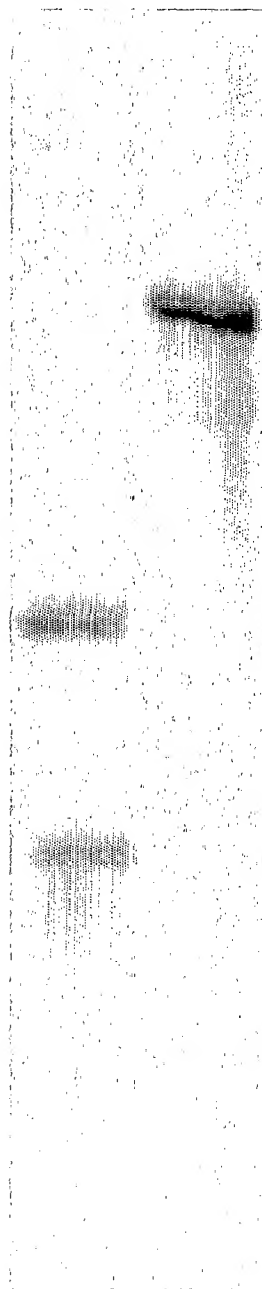
Figure 20.



34/60

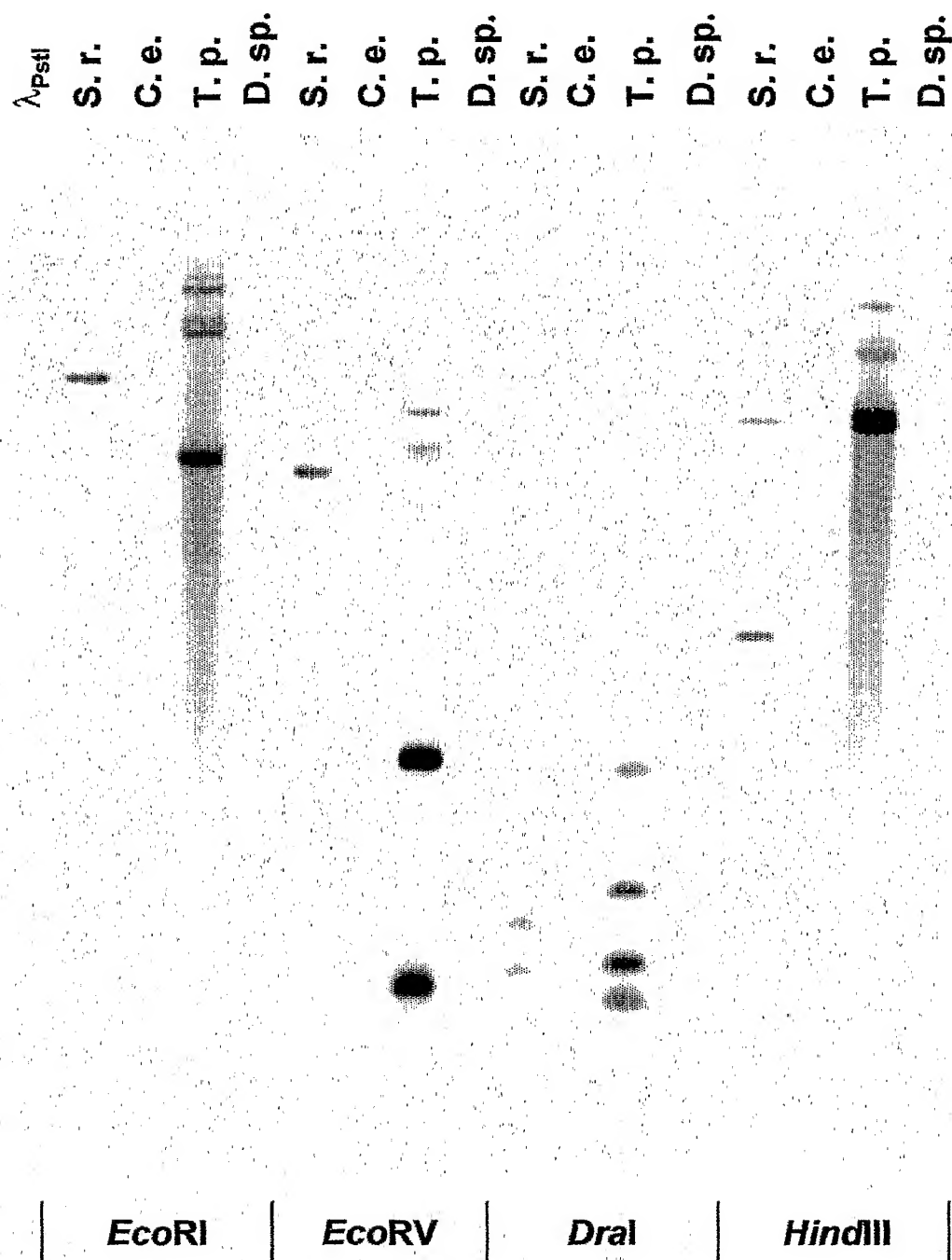
Figure 21.

35/60

Figure 22.*EcoRI* *EcoRV*

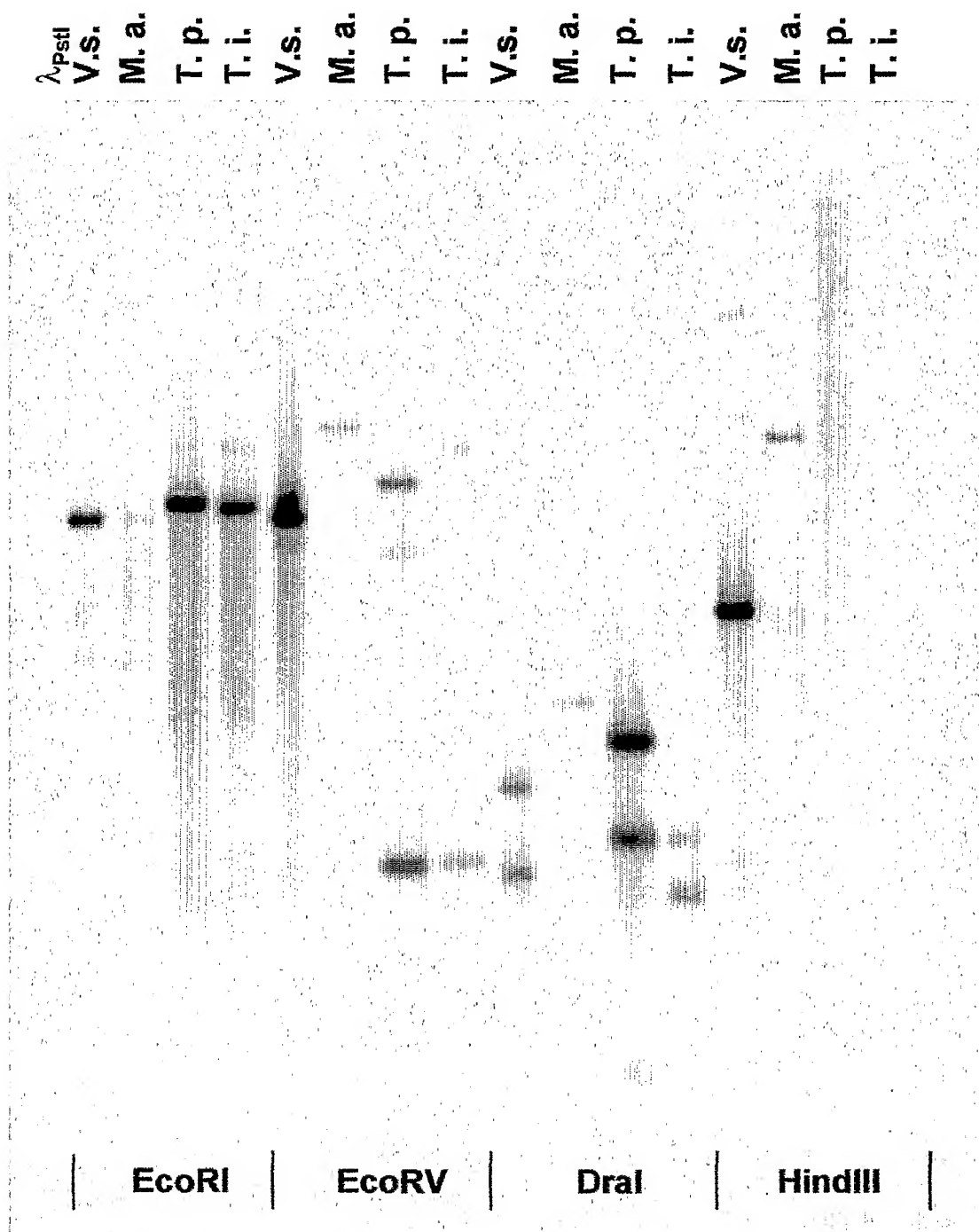
36/60

Figure 23.



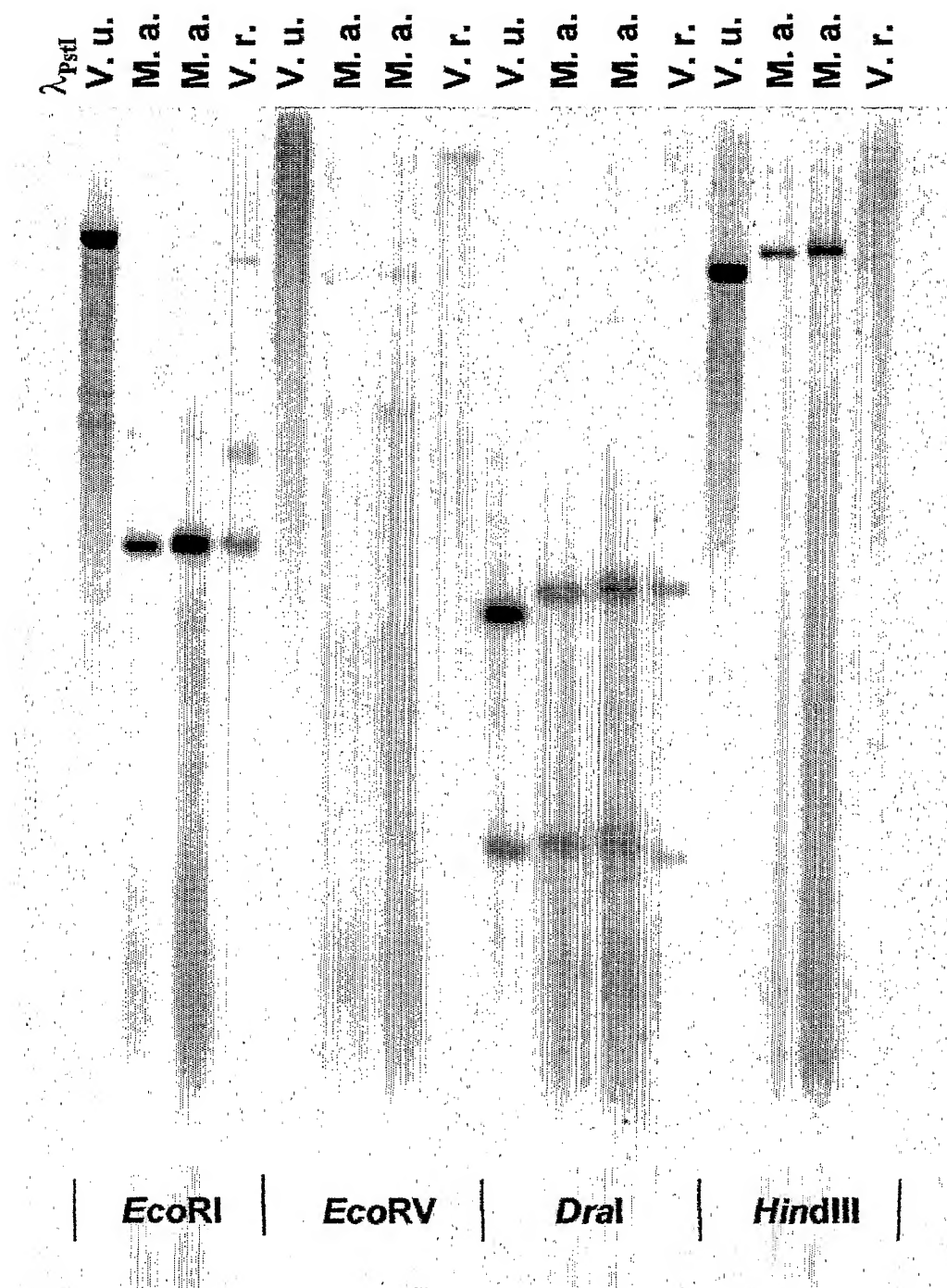
37/60

Figure 24.



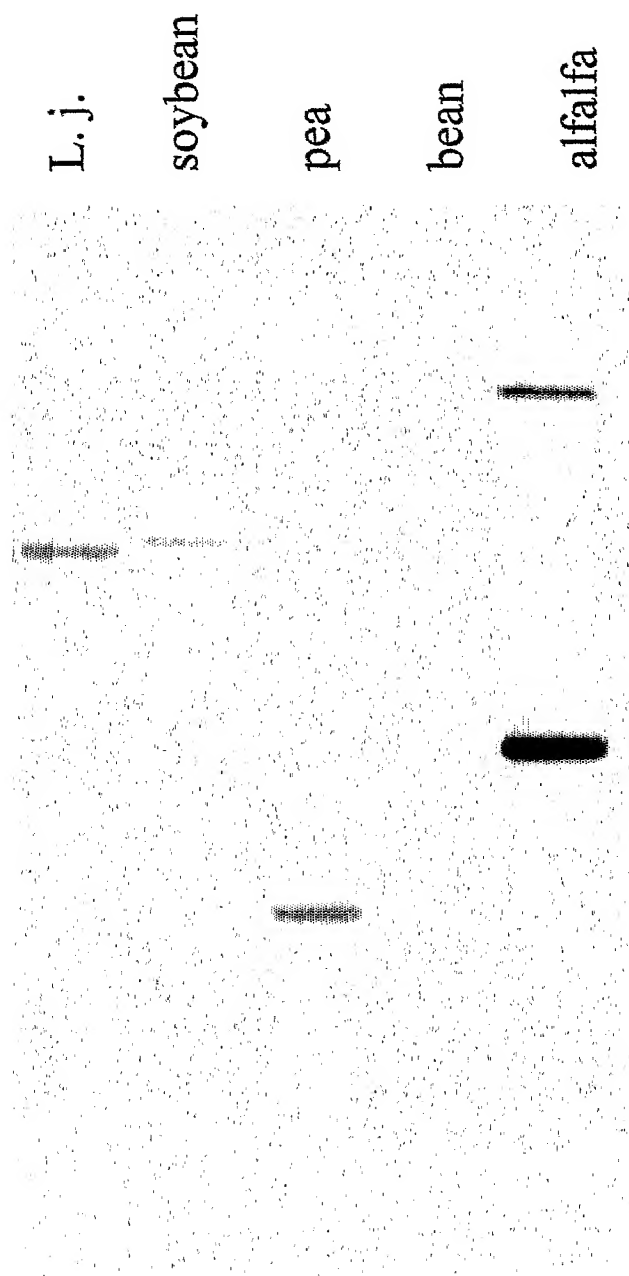
38/60

Figure 25.



39/60

Figure 26.



L. j. = *Lotus japonicus* cv. funakura

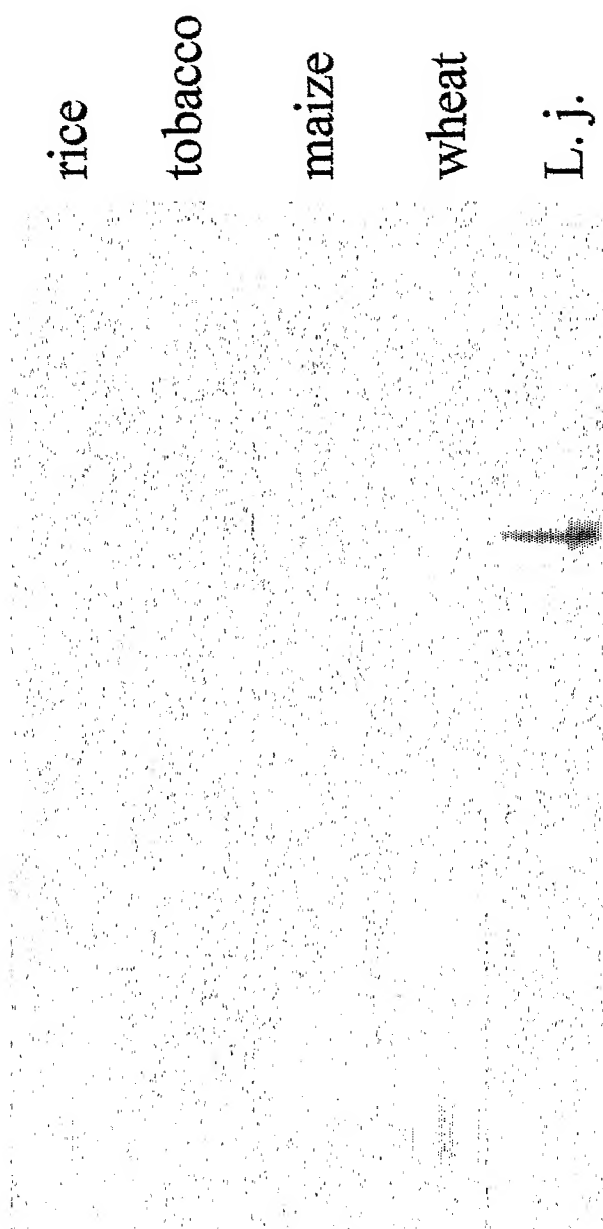
soybean = *Glycine max* cv. Williams

pea = *Pisum sativum* cv. újmajori

bean = *Phaseolus vulgaris* cv. Juliska

alfalfa = *Medicago sativa* ssp. *sativa* (*Eco*RI digestion)

40/60

Figure 27.

rice = *Oryza sativa*

tobacco = *Nicotiana tabacum* cv. Small Havanna SR1

maize = *Zea mays*

wheat = *Triticum aestivum*

L. j. = *Lotus japonicus* cv. funakura

[illegible]

42/60
Figure 29.

[illegible]

43/60

Figure 30.

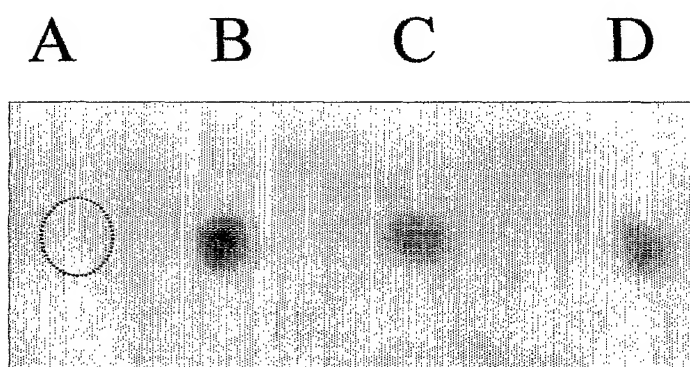
MELRVICIIIRLVVACVLCICI FIRSASSATEGFESIACCADSXXXXXXXXXXXXXXXXXXXXX
XXX
XXX
XXX
XXX
XXX
F
DVLEGGSKYSYTVLNISANGSLNITLVKASGSKFGPLL NAYEILQARPWIDETDQTDLEV
IQKMRKELLQNQDNEALESWSGDP CMLFPWKGVACDGSNGSSVITKLDLSSSNLKG T I P
SSVTEMTKLQILNL SHNHFDGYIPSFPPSSLLISVDLSYNDLTGQLPESII SLPHLNSLY
FGCNQHMRDDDEAKLNSLINTDYGRCAKPKFGQVFMIGAITSGSILITLAVVILFFC
RYRHKSITLEGFGGKTYPMATNII FSLPSKDDFFIKSVSVKPFTLEYIELATEKYKT L I G
EGGFGSVYRGTLDDGQEVAVKVR SATSTQGTREFDNELNLLSAIQHENLVPLLGYCNEYD
QQILVYPFMSNGSLLDRLYGEP AKRKILDWPTRLSIALGAARGLAYLHTFPGRSVIHRDV
KSSNILLDHSMCAKVADFGFSKYAPQEGDSYVSLEVRGTAGYLDPEYYKTQQLSEKSDVF
SFGVVLLEIVSGREPLNIKRPRVEWSLVEWAKPYIRASKVDEIVDPGIKGGYHAEALWRV
XXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXX

44/60

Figure 31.

XX
XX
XX
XX
XXXXXXXXGIDFCLLKEDANPFISQLELRPLPEEYMHDFSTSVLKLINRNNLCGIEDDIRF
PVDQNDRIWKATSTPSYAVPLSFNVSDVDLNGKVTPPRQVLQTALTHPDRLVFVHDGLET
DDYEYSVLLYFLELNNTVKAGQRVFDIYLNSEIKKESFDVSEGGSKYSYITLNIANGSL
NITLAKASGSKFGPILLNAYEILQARPWIDETDQTDVEVIQKLKELLLQNQDNEALESWS
GDPCMLFPWKGVACDGSNGSSVITKLDLSSSNLKGTPSSVTEMTKLQILNLSHNHFDGY
IPSFPSSSLISVDLSYNDLTGQLPESIISLPHLNSLYFGCNQHMSDDDEAKLNSSLIST
DYGRCKAKSPKFGQVFMIGAITSGSILITLAVGILFFCRYRHKSITLEGGGKTYPMATN
IIFSLPSKDDFFIKSVSVKPFSLLEYIELATEKYKTLIGEGGFGSVYRGTLDDGQEVAVKV
RSATSTQGTREFDNELNLLSAIQHENLVPLLGYCNEYDQQILVYPFMSNGSLLDRLYGEP
AKRKILDWPTRLSIALGAARGLAYLHTFPGRSVIHRDVKSSNILLDHSMCAKVADFGFSK
YAPQEGDSYVSLEVRGTAGYLDPEYYKTQQLSEKSDVFSFGVVLEIVSGREPLNIKRPR
VEWSLVEWAKPYIRASKVDEIVDPGIKGGYHAEALWRVVEVALQCLXXXXXXXXXXXXXX
XX

45/60

Figure 32.

46/60

Table 1.

Name of the F2 families	Number of the F2 seeds	Number of the germinated F2 seeds	Number of the died F2 plants	Number of the Nod ⁻ F2 plants	Number of the Nod ⁺ F2 plants	Nod ⁻ : Nod ⁺ ratio
NAB	5571	4988	2412	50	2526	1 : 50
NBW	1285	1199	460	13	726	1 : 55

47/60

Table 2.

F2 → individuals	N	B	W	6	20	37	162	210	790	708	717	11	36	177	743	767	9	23	298	1	2	3	4	7	13	14	19	21	30	42	74	206	243
Name of ↓ markers	N	B	W	6	20	37	162	210	790	708	717	11	36	177	743	767	9	23	298	1	2	3	4	7	13	14	19	21	30	42	74	206	243
OPW8a																																	
OPE8c																																	
nn1																																	
OPB13b																																	
OPA6a																																	

F2 → individuals	N	A	B	267	637	646	765	803	809	871	897	968	1241	908	4156	65	701	4443	334	1219	643	300	12	23	24	82	235	438	782	814	939	1166	1260
Name of ↓ markers	N	A	B	267	637	646	765	803	809	871	897	968	1241	908	4156	65	701	4443	334	1219	643	300	12	23	24	82	235	438	782	814	939	1166	1260
OPW8a																																	
OPE8c																																	
nn1																																	
OPB13b																																	
OPA6a																																	

Table 4.

F2 individuals Name of the markers	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	N A B	
	287	637	646	765	803	809	871	897	968	1241	908	4156	65	701	4443	334	1219	643	300	12	23	24	82	235	438	782	814	939	1166	1280			
OPW8a		3		3	3			3	3	3	5	5	5	5	5	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
U71	3	3	3	3	3	3	3	3	3	3	3		3	3		5		5	5			5				5	5	5	5	5	5	5	
OPE8c	3	3	3	3	3	3	3	3	3	3	3		3	3	3	5	5	5	5	5	5	5		5	5	5	5			5	5	5	
U224		3										5	3	3	3				5							5	5			5	5	5	
CG13	3								3	3	3	5	3	3	3						5		5			5	5	5	5	5	5	5	5
SHMT									3	3	3			3	3		5	5	5	5	5	5	5			5	5			5	5	5	5
nn1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Q5E	3	3	3									3	5	3	5		5	5	5	5	5	5		5		5	5	5	5	5	5	5	5
U492	3	3	3	3	3	3	3	3	3	3	3	3	5	5	5	5	5			5	5		5	5	5	3	5	5	5	5	5	5	5
OPB13b	3	3	3	3	3	3	3	3	5	5	3	3	5	5	5	5	5	3	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5

50/60

Table 5. (1/4)

28I12

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G18P1	Hind III	1,9
G18P4	Hind III	3,8
G18P9	Hind III	3,6
G18P17	Hind III	4,0
G18P24	Hind III	1,0
G18P31	XbaI	3,4
G18P32	XbaI	3,8
G18P37	XbaI	2,4
G18P65	EcoRI	1,0
G18P71	EcoRI	2,8
G18P75	BamHI	1,4

50E23

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G13P1	random	>0,5
G13P2	random	>0,5
G13P3	random	>0,7
G13P4	random	>0,8
G13P5	random	>0,7
G13P6	random	>0,7
G13P7	random	>1,0
G13P8	random	1,5
G13P9	random	1,3
G13P10	random	0,9
G13P11	random	1,5
G13P16	random	0,8
G13P32	random	1,5
G13P33	random	>0,4
G13P41	random	1,6
G13P42	random	1,3
G13P45	random	1,0
G13P47	random	1,0
G13P49	random	2,6
G13P50	random	>0,5
G13P51	random	>0,6
G13P53	random	1,3
G13P55	random	1,4
G13P58	random	1,0
G13P59	random	>1,0
G13P60	random	>0,5
G13P61	random	1,3
G13P62	random	1,1
G13P64	random	1,3

50E23

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G13P65	random	1,5
G13P66	random	1,6
G13P68	random	1,6
G13P69	random	>0,5
G13P70	random	1,4
G13P71	random	1,4
G13P72	random	1,4
G13P73	random	1,5
G13P74	random	>0,6
G13P75	random	1,2
G13P76	random	1,6
G13P77	random	1,6
G13P78	random	1,7
G13P79	random	1,6
G13P80	random	>0,6
G13P81	random	>0,5
G13P82	random	1,3
G13P83	random	1,2
G13P84	random	>0,1
G13P85	random	1,2
G13P86	random	>0,7
G13P87	random	1,3
G13P88	random	>0,6
G13P89	random	>0,8
G13P92	random	>0,8
G13P94	random	1,2
G13P95	random	>0,3
G13P96	random	>0,8
G13P97	random	0,8

51/60

Table 5. (cont. 2/4)

50E23

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G13P98	random	>0,6
G13P99	random	1,4
G13P100	random	0,2
G13P101	random	0,5
G13P104	random	1,5
G13P105	random	6,0
G13P106	random	1,4
G13P107	random	0,7
G13P108	random	6,1
G13P110	random	0,8
G13P113	random	0,1
G13P114	random	0,3
G13P115	random	0,1
G13P116	random	0,7
G13P117	random	0,7
G13P119	random	1,5
G13P120	random	0,8
G13P121	random	1,2
G13P122	random	1,6
G13P123	random	0,8
G13P126	random	1,5
G13P127	random	1,5
G13P128	random	1,0
G13P134	random	1,3
G13P138	random	1,1
G13P153	random	1,3
G13P156	random	1,1
G13P157	random	1,2
G13P158	random	1,7

50E23

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G13P159	random	>0,5
G13P161	random	1,3
G13P165	random	>0,6
G13P166	random	1,5
G13P167	random	1,3
G13P168	random	0,6
G13P169	random	>0,9
G14P1	Cla deléció	3,8

2D11

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G3P17	random	1,6
G3P46	random	0,8
G3P71	random	1,2
G3P73	random	1,2
G3P74	random	1,7
G3P91	random	1,5
G3P92	random	0,9
G3P99	random	>1,7
G3P105	random	1,7
G3P107	random	2,3
G3P123	random	1,3
G3P126	random	1,5
G3P136	random	1,0
G3P139	random	1,3
G3P147	random	1,1
G3P168	random	1,4
G3P178	random	1,5
G3P207	random	0,9
G3P222	random	1,1
G3P224	random	1,2
G3P239	random	1,1
G3P245	random	0,6
G3P268	random	0,8
G3P287	random	0,3
G3P289	random	0,9
G3P301	random	0,9
G3P411	random	1,1
G3P413	random	0,6
G3P477	random	0,9

52/60

Table 5. (cont. 3/4)

2D11

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G3P483	random	>1,7
G3P499	random	0,9
G3P512	random	>0,5
G3P514	random	>0,7
G3P555	random	0,7
G3P576	random	1,2
G3P578	random	0,7
G5P1	BamHI	1,5
G5P4	BamHI	1,0
G5P10	BamHI	3,1
G5P14	BamHI	3,2
G6P3	Hind III	1,7
G6P4	Hind III	1,6
G6P6	Hind III	2,7
G6P11	Hind III	0,8
G6P14	Hind III	1,7
G6P15	Hind III	0,5
G6P17	Hind III	1,8
G6P18	Hind III	1,6
G6P21	Hind III	0,3
G6P26	Hind III	1,8
G10P4	Hind III	2,7
G10P8	Hind III	2,7?
G10P25	Hind III	4,0
G11P1	Hind III	3,3
G11P9	Hind III	9,1?
G11P13	Hind III	9,0
G11P17	Hind III	6,0
G11P20	Hind III	3,3
G11P24	Hind III	8,0

2D11

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G11P34	Hind III	4,4
G11P36	Hind III	7,4
G11P37	Hind III	0,6
G11P38	Hind III	4,7
G11P45	Hind III	2,5
G19P21	Taq deletion	0,9
G19P27	Taq deletion	0,6
G19P42	Taq deletion	0,1
G19P46	Taq deletion	0,8
G19P48	Taq deletion	0,8
G19P73	Taq deletion	0,8
G19P74	Taq deletion	0,7
G19P81	Taq deletion	0,6
G19P94	Taq deletion	0,8

67A11

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G9P13	random	0,8
G9P43	random	1,1
G9P62	random	1,0
G20P1	NheI	0,9
G20P2	NheI	9,6
G20P4	NheI	6,5
G20P5	NheI	8,8
G20P7	NheI	4,4
G20P14	NheI	1,4
G20P15	NheI	5,6
G20P21	NheI	0,4
G20P31	NheI	5,0
G20P38	NheI	1,7
G20P56	NheI	12,0
G20P140	NheI	5,5
G21P25	XbaI	8,6
G21P26	XbaI	4,2
G21P34	XbaI	6,5
G21P35	XbaI	2,2
G21P43	XbaI	3,3
G21P90	XbaI	2,0
G21P136	XbaI	2,6
G21P186	XbaI	6,5
G21P319	XbaI	5
G21P448	XbaI	3,5
G22P2	PaeI-TaqI del.	3,8
G22P5	PaeI-TaqI del.	0,9
G22P7	PaeI-TaqI del.	4,1
G22P8	PaeI-TaqI del.	2,1
G22P19	PaeI-TaqI del.	4,5

53/60

67A11

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G23P4	SacI-TaqI del.	6,1
G23P5	SacI-TaqI del.	3,3
G23P17	SacI-TaqI del.	7,2
G23P23	SacI-TaqI del.	2,6
G24P245	random	1,0
G24P413	random	0,9
G24P418	random	1,1
G24P442	random	1,1
G29P10	Hin2I	1,5
G29P23	Hin2I	0,7
G29P96	Hin2I	4,5
G29P117	Hin2I	0,8
G29P141	Hin2I	0,9
G30P193	Hin6I	6,3
G31P81	TaqI	5,5
G32P68	TaqI	0,4
G32P90	TaqI	0,3
G32P96	TaqI	1,8
G32P112	TaqI	2,7
G32P126	TaqI	5,0
G32P194	TaqI	0,5
G32P286	TaqI	3,3
G33P16	EcoRI	6,0
G33P48	EcoRI	2,3
G33P63	EcoRI	5,0
G33P64	EcoRI	3,0
G33P121	EcoRI	4,5
G33P155	EcoRI	1,6
G34P20	HindIII	1,4
G34P30	HindIII	7
G34P34	HindIII	2,3

Table 5. (cont. 4/4)

67A11

NAME OF THE CLONE	RESTRICTION ENDONUCLEASE	Insert size (kb)
G34P42	HindIII	1,8
G34P44	HindIII	4,2
G34P46	HindIII	2,8
G34P52	HindIII	4,6
G34P64	HindIII	2,6
G34P101	HindIII	0,6
G34P173	HindIII	2,2
G34P192	HindIII	3,0
G35P65	SauIII	0,7
G35P85	SauIII	1,4
G35P136	SauIII	1,3
G50P4	KpnI deletion	3,5

55/60

Table 6.

Structural element	First nucleotide	Last nucleotide
	Position	
EXON 1	<897	1248
EXON 2	1564	2074
EXON 3	2798	3273
EXON 4	3372	3522
EXON 5	3800	3871
EXON 6	3979	4047
EXON 7	4519	4590
EXON 8	4702	4773
EXON 9	4860	5032
EXON 10	5542	5768
EXON 11	5912	6037
EXON 12	6110	6178
EXON 13	6797	6986
EXON 14	7340	7472
EXON 15	7562	8236<

INTRON 1	1249	1563
INTRON 2	2075	2797
INTRON 3	3274	3371
INTRON 4	3523	3799
INTRON 5	3872	3978
INTRON 6	4048	4518
INTRON 7	4591	4701
INTRON 8	4774	4859
INTRON 9	5033	5541
INTRON 10	5769	5911
INTRON 11	6038	6109
INTRON 12	6179	6796
INTRON 13	6987	7339
INTRON 14	7473	7561

56/60

Table 7.

NAME OF THE CLONE	CLONING VECTOR	ENZYMES	MODIFICATION	NORK FRAGMENT	ANTIBIOTIC RESISTANCE (µg/ml)	Name of the <i>E. coli</i> strain
pBRC1667	pAT680	<i>Kpn</i> I	dephosphorilation	8.8 kb <i>Nhe</i> I	kanamycin (200)	BRC1667
pBRC1678	pPK459	<i>Kpn</i> I	dephosphorilation	8.8 kb <i>Nhe</i> I	tetracycline (10) kanamycin (50)	BRC1678
pBRC1666	pPR97	<i>Cla</i> I-Sac I	dephosphorilation	8.8 kb <i>Nhe</i> I	kanamycin (200)	BRC1666
pBRC1701	pPR97	<i>Cla</i> I	dephosphorilation	11.8 kb <i>Cla</i> I	kanamycin (200)	BRC1701

57/60

Table 8.

Agrobacterium strain	TRANSFORMED PLASMID	NAME OF THE STRAIN CARRYING THE NORK GENE
<i>A. rhizogenes</i> A4	pBRC1667	BRC1673
<i>A. rhizogenes</i> 1334	pBRC1667	BRC1675
<i>A. rhizogenes</i> 1724	pBRC1667	BRC1679
<i>A. tumefaciens</i> LBA4404	pBRC1667	BRC1677
<i>A. tumefaciens</i> LBA4404	pBRC1678	BRC1680
<i>A. tumefaciens</i> LBA4404	pBRC1666	BRC1681
<i>A. tumefaciens</i> LBA4404	pBRC1701	BRC1707

58/60

Table 9.

NAME OF THE STRAINS CARRYING THE NORK GENE	NUMBER OF THE PLANTS DEVELOPING ROOTS	NUMBER OF THE GUS + ROOTS	NUMBER OF THE NOD + PLANTS
BRC1673	78 (5)	1 (0)	0
BRC1675	87 (2)	3 (0)	0

59/60

Table 10.

Plants to be transformed	Number of shoots or leaves treated with Agrobacterium	Number of regenerated plants	Number of transformed plants	Number of Nod ⁺ plants
<i>F1RN28</i>	157	19	2	2
<i>F1RN41</i>	133	14		
F2RN28/4	120			
F2RN28/5	75			

60/60

Table 11.

Plants to be transformed	Number of shoots or leaves treated with Agrobacterium	Total number of transformants	Number of independent transformants/ Total number of transformants	Presence of the M. truncatula NORK sequence/Number of independent transformants	Presence of the NORK protein/Number of independent transformants
<i>Nicotiana tabacum</i>	113	15	6/15	6/6	6/6
<i>Vicia/Lotus</i>	232	26	8/26	8/8	nt